

Christian Helmers and Henry G. Overman
**My precious! The location and diffusion of
scientific research: evidence from the
Synchrotron Diamond Light Source**

**Article (Published version)
(Refereed)**

Original citation:

Helmers, Christian and Overman, Henry G. (2017) *My precious! The location and diffusion of scientific research: evidence from the Synchrotron Diamond Light Source*. The Economic Journal . ISSN 0013-0133

DOI: [10.1111/econj.12387](https://doi.org/10.1111/econj.12387)

Reuse of this item is permitted through licensing under the Creative Commons:

© 2016 The Authors. The Economic Journal published by John Wiley & Sons Ltd on behalf of Royal Economic Society
CC BY 4.0

This version available at: <http://eprints.lse.ac.uk/64659/>

Available in LSE Research Online: April 2017

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

MY PRECIOUS! THE LOCATION AND DIFFUSION OF SCIENTIFIC RESEARCH: EVIDENCE FROM THE SYNCHROTRON DIAMOND LIGHT SOURCE

Christian Helmers and Henry G. Overman

We analyse the impact of the establishment of the Diamond Light Source synchrotron, a large basic scientific research facility in the UK, on the geographic distribution of related research. To account for the potentially endogenous location choice of the synchrotron, we rely on the availability of a 'runner-up' site. We use data on academic publications to trace the geographic distribution of related scientific inputs and outputs. Our results suggest that proximity to Diamond had a positive impact on the output of related research. This proximity effect appears to be driven by an increase in inputs rather than the productivity of scientists.

We investigate the impact of a £380 million scientific research facility on the geographic distribution of the knowledge created by the facility. The Diamond Light Source, a 3rd generation synchrotron light source, represents the single largest investment in research infrastructure in the modern history of the UK. The facility, which opened in January 2007, is one of only a few 3rd-generation synchrotron facilities worldwide. It enables researchers to conduct novel scientific experiments that are likely to shift the knowledge frontier in a number of scientific disciplines.

We are primarily interested in whether the location choice of Diamond has affected the geographic distribution of scientific research in relevant scientific fields. The research question that we address is whether the establishment of basic scientific research infrastructure, which is inherently indivisible, disproportionately benefits institutions (and hence areas) that are closer to the infrastructure. This under-researched question is particularly relevant with regard to 'lumpy', large-scale infrastructure investments. The analysis also sheds light on the formation of research clusters and on the geographical distribution of scientific research.

The first challenge in establishing the impact of Diamond is due to the lumpiness of the investment: Diamond is based on a single site. However, as our interest is in the geographic effects of Diamond, we assume that all institutions can benefit from

* Corresponding author: Christian Helmers, Economics Department, Santa Clara University, Lucas Hall, 500 El Camino Real, Santa Clara, CA 95053, USA. Email: chelmehrs@scu.edu.

This work was supported by the Economic and Social Research Council grant numbers ES/J021342/1 and ES/G005966/1. We thank Zhe Sun and Wenjie Wu for excellent research assistance and Felix Weinhardt for assistance with GIS. We thank Manos Kitsios, Tim Simcoe, seminar/session participants at UC Berkeley, UC Merced, EPFL Lausanne, KU Leuven, Stanford, Santa Clara, the University of Würzburg, the University of Kassel, the SERC Conference 2011 at LSE, the Royal Economic Society Conference 2013, the NBER Innovation Summer Institute, a workshop at Universitat de Barcelona, the Technology Transfer Conference 2011 and the 7th Meeting of the Urban Economics Association for their useful comments and suggestions. We are particularly grateful to Walter Luyten for advice on the data construction. Finally, we thank the editor Kjell Salvanes and two anonymous referees for comments and suggestions that have considerably improved the article.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Diamond and then test whether the extent of any benefit decays with distance from Diamond. That is, we test whether the intensity of treatment depends on distance to Diamond. This approach has been used by previous studies using historical ‘experiments’ to look at the geographic effects of large lumpy shocks, see, for example Hanson (1997) on the effect of NAFTA, Draca *et al.* (2011) on the impact of the terrorist attacks on London and Ahlfeldt *et al.* (2012) on the impact of the Berlin Wall.

In common with that literature, a second challenge in establishing a causal effect of proximity to Diamond on scientific research arises because other factors that affect scientific research may also differ in their proximity to Diamond. If the location of the investment is chosen to be close to institutions that are strong in the relevant scientific fields, then looking at the effect of distance to Diamond will overstate the benefits of proximity. Controlling for observable characteristics of institutions, e.g. using measures of their strength in relevant fields, helps address these concerns. But this leaves open the possibility that unobservable characteristics of institutions determined the choice of location; in that event, a positive effect of proximity may reflect the fact that these unobserved characteristics vary with proximity to the chosen location. In the case of the Diamond Light Source, we partially address this problem by exploiting the availability of a ‘control’ location. Diamond was built at the Harwell Science and Innovation Campus at the Rutherford Appleton Laboratory in Didcot, Oxfordshire. But there had previously been concrete plans to locate Diamond about 215 kilometres away in another research hub in Daresbury near Manchester. The final decision to locate Diamond in Oxfordshire was preceded by a heated political debate over the siting. If we assume that the same unobservables determined the choice of both the final and the runner-up location, then comparing the effects of proximity to these two alternative locations helps control for the impact of those unobservables.

If this identifying assumption fails, then positive unobserved shocks correlated with distance to the final location (or negative shocks correlated with distance to the runner-up location) could drive our results, even if proximity to Diamond has no effect on scientific research. The availability of multiple winners and runners-up would help mitigate concerns about confounding unobservables that are correlated with distance to either location (see, in different settings, Greenstone *et al.*, 2010; Malmendier *et al.*, 2012). But for large lumpy investments of the kind we consider here, estimation requires the maintained assumption of similarity between the final and runner-up locations. Nevertheless, this assumption is weaker than that used in the large distance decay literature which, in the absence of a counterfactual location, simply assumes that unobservables are uncorrelated with distance to the final location of investments.

Our main focus is on tracing the geographical distribution of relevant research using scientific publications. This codified form of knowledge is a particularly suitable measure of research output in our analysis given the nature of the scientific research enabled by Diamond. Research conducted at Diamond can be regarded as ‘cutting edge’, which makes it likely findings will be publishable in scientific academic journals. Second, research at Diamond focuses on highly codifiable scientific disciplines such as structural biology, physics, chemistry and materials science – fields in which output is routinely published in scientific journals and where the publication lag is often negligible.

Our findings suggest that distance to Diamond had a significant effect on the extent to which research in related fields benefited from the investment in Diamond.

Specifically, in our main results, we find a statistically significant and economically important increase in our measure of scientific output of around four standard deviations within a 25 kilometre radius of Diamond. This effect applies to research that is generated from experiments carried out directly at Diamond as well as related research that does not use the synchrotron facility. Comparison to Daresbury allows us to interpret this effect as causal: proximity to Diamond causes increases in output for locations close to Diamond over and above those that would have occurred if Diamond had been located elsewhere. That said, our strategy does not allow us to consider the overall effect of Diamond, so all locations may benefit from the investment, but our results do show that institutions (and hence areas) close to Diamond benefit significantly more than those further away. We demonstrate that these results are robust to a number of variations of our empirical specification, the choice of counterfactual location, as well as changes in the construction of our measures of research input and output.

We consider a number of possible explanations of this effect on research outputs. The explanations split into two groups: either the effect results from an increase in inputs or from an increase in productivity (or a combination of both). Looking at the article by the author counts suggests that the effect is not explained by changes in productivity. However, we do see an increase in inputs, specifically the number of scientists working on relevant research in proximity to Diamond. When we look at the movement of authors across locations within the UK we see that, while authors on average do move closer to Diamond post-opening, the net change very close to Diamond is zero. At the same time, we find an increase in the number of authors that move close to Diamond from abroad directly after Diamond opens – although the number of movers is not sufficient to explain our main result. This suggests that the direct and indirect effects at the extensive margin are driven by ‘new’ scientists (rather than relocations). We present results (using a database of crystallographers) that suggest that this effect at the extensive margin is driven mostly by an increase in the number of scientists located close to Diamond. For knowledge production, we argue that co-authoring provides an additional mechanism for increasing inputs – although we show that this kind of indirect increase in inputs cannot explain our results.

Our results contribute to the literature on the importance of agglomeration externalities produced by indivisible scientific research facilities for science and innovation. This literature focuses overwhelmingly on externalities between companies (Jaffe *et al.*, 1993; Audretsch and Feldman, 1996*a*) or from university to private industry (Jaffe, 1989; Kantor and Whalley, 2014*a,b*). We offer for the first time empirical evidence on the importance of local externalities created by basic scientific research infrastructure in forming clusters of scientific research.

This article is organised as follows. Section 1 provides detailed information on Diamond and its location choice. Section 2 outlines the empirical approach taken to identify the impact of the establishment of Diamond. Section 3 describes the data used in the analysis while Section 4 discusses the results. Section 5 presents a number of robustness checks and Section 6 investigates the mechanisms behind the effects found in our analysis. Section 7 concludes.

1. Diamond Light Source

1.1. UK's 3rd Generation Synchrotron

The Diamond Light Source is a synchrotron facility. Synchrotron facilities are circular particle accelerators that produce beams of X-rays, infrared and ultraviolet light (see Figure 1).¹ Such synchrotron light is useful to study small objects, such as molecules and atoms, whose visualisation requires light with shorter wavelengths than available in microscopes.² Diamond consists of a 561m storage ring and has a total floor area of 45,500m².

Diamond is funded by the UK Science and Technology Facilities Council (STFC) (86%) and the Wellcome Trust (14%). After the location decision had been taken in March 2000, the two-phase construction of Diamond started in early 2003. In Phase I, the buildings for the synchrotron facility were constructed and the first seven beamlines established (at a cost of £263 million). User operations on the Phase I beamlines began in January 2007. In Phase II, another 15 beamlines are being added to the facility, requiring a further £120 million in investment. There are currently 19 operational beamlines which are used to conduct experiments in various fields including condensed matter physics, materials science, biology and medicine serving both basic and applied research. This provides ample scope for the creation of new publishable knowledge in a range of applied scientific fields.

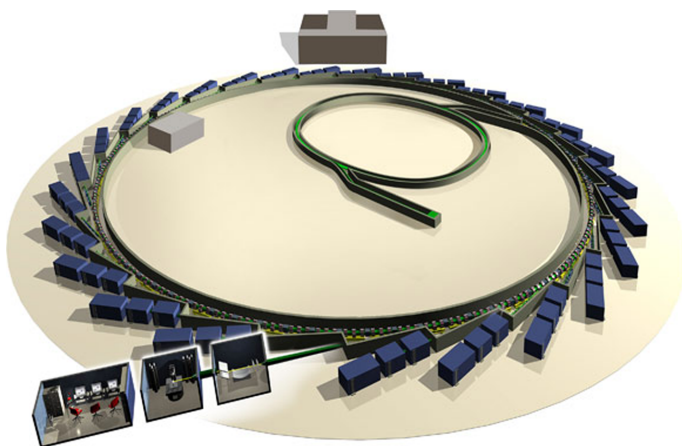


Fig. 1. *Diamond Synchrotron Facility*

Source. Diamond Light Source.

¹ A synchrotron consists of a large ring-shaped tube which accelerates charged particles fired into it from a linear accelerator. The ring is enclosed by magnets that keep the particles 'on orbit'. The accelerated particles are ejected into a 'storage ring' in which they circulate. The continuous movement of the electrons, which is created by 'insertion devices', results in electromagnetic waves ('synchrotron radiation'). This radiation is captured in beamlines which use the radiation for experiments.

² There are three types of experiments that can be conducted at Diamond:

- (i) X-ray diffraction and scattering;
- (ii) spectroscopy; and
- (iii) imaging and microscopy.

Diamond superseded the existing UK synchrotron which was located at the STFC Daresbury Laboratory near Manchester. The Synchrotron Radiation Source (SRS), opened in 1981, was the second UK synchrotron light source. It replaced the UK's first synchrotron NINA built in 1964 also in Daresbury (NINA was closed in 1977). Given Diamond's technical superiority, the SRS became obsolete and was closed in August 2008. The main difference between a 2nd generation synchrotron and a 3rd generation synchrotron is how synchrotron light is generated. While 2nd generation synchrotrons (SRS) rely on dipole bending magnets to produce synchrotron radiation, 3rd generation synchrotrons (Diamond) rely on the so-called undulators/wigglers which cause electrons to wiggle producing more intense, brighter synchrotron light. This allows higher resolution and improves the synchrotron's applicability for X-ray microscopy to spectromicroscopy which benefits particular scientific fields such as crystallography.

Beamtime is granted after submission of a proposal which specifies the amount of time the research team would like to use the facility and the beamline that will be used. Beamtime is allocated to academic users through a scientific peer review panel and a panel that assesses technical feasibility.³ Beamtime is free for academic users and corporate users that commit to putting the research results into the public domain. Private companies wanting to maintain the ownership of any intellectual property resulting from their work at Diamond may apply for beamtime but are liable to a usage fee.⁴

1.2. *Location Choice*

Our primary identification strategy rests on a controversy that arose in the siting of Diamond. Initially, the government had firm plans to site the new synchrotron at the STFC Daresbury Laboratory next to the existing UK synchrotron. However, the Wellcome Trust suggested that the new synchrotron should be built instead at the Harwell Science and Innovation Campus in Didcot (Oxfordshire) effectively co-locating Diamond with the Rutherford Appleton Laboratory (RAL). According to news articles, Wellcome believed that 'greater scientific benefits would result from a location close to the existing neutron source [ISIS] and to Medical Research Council units and the University of Oxford' (Loder, 1999). Hence, the main argument was to concentrate research facilities in a single location (Didcot was already home to ISIS, one of the leading pulsed neutron and muon sources worldwide) to strengthen national centres of excellence in research. Supporters of the Daresbury location, in contrast, argued that given the expected applicability of Diamond to only a limited number of scientific disciplines, Wellcome was overstating the importance of geographical proximity to the so-called Oxford-London-Cambridge Golden Triangle. Instead, they argued that relocation of staff from the Daresbury to Didcot would

³ The members of the academic peer review panel come from a wide range of UK and European universities, research institutions, and private companies. The list of panel members is available on Diamond's website: <http://www.diamond.ac.uk/Home/ForUsers/academics/panel.html>

⁴ Of the beamtime, 80% is allocated to external, i.e. academic and industrial, users. Industrial users can only use up to 10% of the beamtime for external users.

represent a substantial but unnecessary expense and deprive the Manchester region of publicly funded top scientists employed at the SRS. The controversy received broad public attention and led to heated debates in Parliament as well as to discussions in a large number of news channels and newspapers including reports by the *BBC*, *Financial Times*, the *Times Higher Education* and *The Guardian*, as well as scientific media, such as *Nature*. The issue received particular public interest as supporters of the Daresbury site framed the controversy within the longstanding debate on the North–South divide in terms of scientific research infrastructure in the UK.

In March 2000, the government announced that the synchrotron would be built at the Rutherford Appleton Laboratory near Oxford. However, the debate continued and even more than a decade later, Diamond's website still justifies this decision by stating that '[t]he Harwell Science and Innovation Campus is a thriving hub of scientific research and there is a high concentration of users within the region. Diamond is surrounded by a number of scientific research facilities making the site a centre of excellence in terms of tools and expertise and therefore the ideal location for the UK's new synchrotron'. This statement implies that geographical proximity to potential users is the main argument in favour of the decision to locate the facility in Didcot near Oxford. The underlying assumption is that geographical proximity influences not only a potential user's decision to employ the facility to conduct research but also the impact of the resulting scientific output. However, the debate surrounding the decision to locate Diamond near Oxford and the arguments offered by both sides suggest that from a scientific point of view both locations were *ex ante* similarly competitive clusters with respect to research that could be conducted using a synchrotron. This provides the basis for our identification strategy outlined in the following Section.

2. Empirical Strategy

2.1. Theoretical Framework

To provide some intuition for the basic mechanisms at work, we start with a simple conceptual framework. We assume that there is a fixed number of geographical areas $a = 1, \dots, N$ in which scientists work in relevant scientific fields. Researchers in area a produce scientific output p_{at} in period t . Output is produced with two inputs:

- (i) research infrastructure S_{at} ; as well as
- (ii) total available research output (i.e. publications) denoted by R_{at} .

We assume that the overall, available research output is a function of total research output excluding the knowledge created in a ,⁵ $R_{at} = g(\sum_{a \neq b} \delta_{ab} p_{bt})$ where δ_{ab} is a constant that reflects the geographical proximity between areas a and b (e.g. inverse geographic distance). The fact that research infrastructure S_{at} is area-specific captures the notion that access to a research facility is easier the closer the facility is located to a researcher. This assumption still allows for an infrastructure shock to affect several

⁵ Including p_{at} would produce a trivial feedback effect since we assume for simplicity that these effects are contemporaneous.

areas. In addition, the production function includes a productivity shifter A_{at} specific to area a in period t which combines a range of factors such as area-specific technology shocks which affect the productivity of the two inputs equally. With these assumptions in hand, we write a simple production function as follows:

$$p_{at} = f(A_{at}, S_{at}, R_{at}), \quad (1)$$

where p_{at} is increasing in both inputs. This simple framework allows us to illustrate the two main effects at work – a direct and an indirect effect – of an increase in S_{at} as a consequence of Diamond (online Appendix A shows how we derive this expression):

$$\frac{\partial p_{at}}{\partial S_{at}} = \underbrace{\frac{\partial f(A_{at}, S_{at}, R_{at})}{\partial S_{at}}}_{\text{Direct Effect}} \times \underbrace{\left\{ 1 - \left[\frac{\partial f(A_{at}, S_{at}, R_{at})}{\partial R_{at}} \times \Omega \right] \right\}^{-1}}_{\text{Indirect Effect}}, \quad (2)$$

where:

$$\Omega = \frac{\partial g(\cdot)}{\partial p_{at}} \left\{ \sum_{a \neq b} \left[\frac{\partial g(\cdot)}{\partial p_{bt}} \frac{\partial f(\cdot)}{\partial R_{bt}} \right] \right\}.$$

In our set-up, a positive infrastructure shock, i.e. an increase in S_{at} , has a positive direct effect on research output ($[\partial f(A_{at}, S_{at}, R_{at})/\partial S_{at}] > 0$) because it enables researchers to conduct novel experiments that push the knowledge frontier. Hence, even in the absence of externalities, the shock creates more output through this direct effect. In the presence of externalities, however, the shock creates output above and beyond the direct effect. Here, externalities emerge because the output of researchers in area a depends on the aggregate research output. In this way, the infrastructure shock in a directly increases output of researchers in area a (direct effect) and feeds back into the production of output via the total research output, i.e. a 's output affects b 's output which in turn affects a 's output via R_{at} (indirect effect). The indirect effect introduces non-linearity into the production function which produces localised externalities. The effect is localised because R_{at} is increasing in δ_{ab} , that is, the indirect effect increases in geographical proximity between areas a and b . Hence, the indirect effect acts as positive multiplier in (2) as long as $0 < [\partial f(\cdot)]/(\partial R_{at}) \times \Omega < 1$. Obviously, this set-up is simplistic (e.g. it does not explain the optimal choice of inputs) and it should not suggest that local externalities only emerge because of the use of other researchers' output as research input. Still, the framework illustrates one possible way in which geographical proximity to infrastructure affects output directly but also indirectly through its effect on total research output. The following Section explains how we estimate these two effects empirically and test for their presence.

2.2. Empirical Approach

We want to know if the establishment of Diamond in Oxford disproportionately benefited institutions close to Diamond, resulting in the geographical concentration of research and innovative output beyond what would have happened, had Diamond been sited elsewhere. We focus on the geographical distribution of research within the

UK because, as observed by a member of Parliament in a debate on the siting of Diamond '[w]hether one flies from Tokyo to Daresbury or from Tokyo to Oxford is irrelevant.' (A. Stunneel, Hansard, 29 March 2000, col. 63WH).⁶ The object of interest is the geographical distribution of research activity conducted in scientific fields related to Diamond. Our main focus will be on explaining differences in research activity across geographical areas in the UK where we define geographical areas based on travel to work areas (TTWAs).⁷

We rely on observable measures of research for our analysis: academic journal publication and author counts. To investigate potential channels through which the facility contributes to changes in scientific output, we also analyse potential 'byproducts' created by the facility (David *et al.*, 1992), such as the formation of co-authorships, or the entry of 'new' scientists into fields for which the synchrotron is relevant. To help with exposition, in this Section we refer to academic paper counts for TTWAs (the main focus of our empirical analysis). All methods extend readily to alternative counts and different units of observation.

As discussed in the introduction, we face two main challenges in establishing the causal effect of Diamond. First, because Diamond is based on a single site, we must limit our attention to establishing the effect on the geographical distribution of relevant scientific research (rather than the total impact). To do this, we assume that Diamond affects all areas but with the intensity of treatment depending on distance to Diamond. Second, we need to control for the potential endogeneity of Diamond's location which means that unobservable factors that influence scientific research may be correlated with distance to Diamond. As discussed above, our main strategy for dealing with this endogeneity is to exploit the availability of a 'runner-up' location: Daresbury. If we assume that the same unobservables determined the choice of both final and runner-up location, then comparing the effects of proximity to these two alternative locations helps control for the impact of those unobservables.

One concern with this strategy is that Daresbury was affected by the shutdown of the 2nd generation synchrotron which coincided with the opening of Diamond. We address this concern in two ways. First, we estimate the effect of proximity to Diamond relative to the effect of proximity to a third location in the North of England. The third location is also a cluster of relevant scientific research but there were no important changes in scientific infrastructure in its immediate proximity during the period studied. Second, we use the methodology suggested by Abadie and Gardeazabal (2003) to construct a synthetic counterfactual location based on observable characteristics of the location in which Diamond was sited. This means we obtain the geographical effect

⁶ Nevertheless, Section 6 investigates whether scientists who moved to the UK from abroad – potentially as a result of Diamond – contribute to the observed increase in scientists around Diamond. It might also be interesting to investigate other potential international spillovers. For example, British universities and researchers in geographical proximity to Diamond might have found it easier to initiate international collaborations due to the need by foreign researchers to team up with local researchers to have better and more flexible access to the synchrotron facility.

⁷ We face a trade-off in the choice of spatial units for our analysis. Activity is sufficiently 'rare' that we want to aggregate up to avoid problems of excess zeros, but we want to use small spatial scales to better capture any changes to the geographical distribution of activity. For robustness we also provide some results using local authorities (LAs) or universities/research institutions as the unit of analysis.

of Diamond by comparing changes in scientific output as a function of distance to Diamond as compared to distance to:

- (i) the runner-up location, which accounts for cluster-related time-varying unobservables;
- (ii) another science cluster not directly subject to changes in scientific infrastructure; and
- (iii) a synthetic counterfactual which accounts for any endogeneity in Diamond's location choice based on observable characteristics.

The remainder of this Section provides details.

Our starting point is the following estimating equation:

$$p_{at} = \alpha_a + \sum_t \delta_t D_t + \sum_r \gamma_{DI}^r [D_{DI}^r \times I(t \geq 2007)] + \epsilon_{at}, \quad (3)$$

where p_{at} is the count of published academic papers from authors employed in area a at time t ; α_a is a fixed effect for area a ; D_t a dummy variable taking the value one if year is equal to t , zero otherwise; D_{DI}^r are a set of R 'ring' dummies which take value one if the area is within a given distance of Diamond, zero otherwise; $I(t \geq 2007)$ is a 'post-Diamond' indicator variable taking value one from 2007 onwards (the year Diamond opened for external users), zero otherwise; ϵ_{at} is an idiosyncratic error.⁸

In our main analysis, we use three ring dummies corresponding to distances 0–25, 25–125, and 125–175 kilometres.⁹ The ring dummies allow for the fact that the effect of proximity may be non-linear, while the fixed effects control for the fact that research activity would not have been uniformly distributed across areas in the absence of Diamond. In our main analysis, the comparison group comprises areas located more than 175 kilometres from Diamond (the omitted category). In this specification, the interaction of these ring dummies with an indicator for years after the opening of Diamond is intended to capture whether the impact of Diamond varies with proximity to the new facility. The time dummies allow for the fact that aggregate research activity may vary over time. We also include linear pre-trends in all specifications.¹⁰

As usual, anything that causes the error ϵ_{at} to be correlated with the distance to Diamond (as captured by the ring dummies) will bias coefficients on the distance dummies and hence our estimate of the impact of Diamond. The main source of such correlation, in our context, arises because the decision where to locate Diamond was influenced by an assessment of the research potential of different places. The fixed effects capture the role of time-invariant location-specific unobservables that influence both the location decision and affect outcomes. To help further address this problem, we can control for time-varying observable characteristics of locations as follows:

⁸ Note that the level-effect is absorbed as part of the area fixed effect because the distance ring dummies are time-invariant.

⁹ These distance rings use the centroid of the Oxford TTWA as origin. This specification of the distances implies that Oxford is included in the first distance ring, Cambridge and London in the second, and the third ring includes cities such as Nottingham or Cardiff. We verify the robustness of our results for different distance ring definitions in Section 5.

¹⁰ Pre-trends are constructed by interacting the TTWA-fixed effect with the continuous year variable.

$$p_{at} = \alpha_a + \sum_t \delta_t D_t + \sum_r \gamma_{DI}^r [D_{DI}^r \times I(t \geq 2007)] + \beta X_{at} + \epsilon_{at}, \quad (4)$$

where X_{at} are characteristics of areas that may affect research activity. Equation (4) provides consistent estimates of the treatment effect of Diamond if $\text{Cov}(\epsilon_{at}, D_{DI}^r | X_{at}, \alpha_a) = 0 \forall r$. The inclusion of X_{at} controls for the fact that observable area characteristics may drive both the number of papers published and the location of Diamond, introducing correlation between ϵ_{at} and D_{DI}^r in (3). Note, however, that if the location of Diamond causes changes in X_{at} then controlling for it will lead us to underestimate the impact of Diamond.¹¹

In our main results, we include as covariates X_{at} measures of local skill composition (%NVQ4 and above) and the size of the local labour force. To account for the availability of research funding, we also include total grant payments from the STFC and both total grant payments and number of grants from the Wellcome Trust (for details on the construction of these variables see online Appendix B.4 and B.5).

This still leaves the possibility that the endogeneous location decision means there is some unobservable time varying factor correlated with proximity to Diamond. One concrete concern may be the tendency for existing clusters of innovation to strengthen over time (Audretsch and Feldman, 1996*b*; Feldman and Francis, 2004). Given that we know Diamond was sited in an existing research cluster, this will overestimate the impact of Diamond if this clustering effect is observed in the TTWAs that are closer to Diamond (in ways that are not fully captured by observable characteristics X_{at}). We see this as the main identification problem for estimates of the causal effect of Diamond based on (4). As discussed above, to address this concern we use the availability of a runner up location at Daresbury (which also represents an existing cluster of activity in this area) to estimate:

$$\begin{aligned} p_{at} = & \alpha_a + \sum_t \delta_t D_t + \sum_r \gamma_{DI}^r [D_{DI}^r \times I(t \geq 2007)] \\ & + \sum_r \gamma_{DA}^r [D_{DA}^r \times I(t \geq 2007)] + \beta X_{at} + \epsilon_{at}, \end{aligned} \quad (5)$$

where D_{DA}^r are a set of R ‘ring’ dummies which take value one if the area is within a given distance of Daresbury, zero otherwise and everything else is as before.¹² In (5), the effect of being a given distance from Diamond (as defined by the choice of ring dummies) is given by $\theta^r = \gamma_{DI}^r - \gamma_{DA}^r$ where standard errors are obtained using the delta method. That is, in (5) we compare the effect of the opening of Diamond on scientists located within a given distance of Diamond relative to scientists across all other locations to the effect on scientists located within the same distance of Daresbury relative to researchers across all other locations. Because γ_{DI}^r and γ_{DA}^r are obtained from variation in changes of output generated within the r distance rings and all other locations, the identifying assumption is that conditional on unobservable time-invariant and observable time-varying location-specific characteristics, changes in

¹¹ Angrist and Pischke (2009) refer to this as the ‘bad control’ problem.

¹² We correct for overlaps in the Diamond and Daresbury distance rings by allocating a given location to either the Diamond or the Daresbury distance ring depending on whether it is closer to Diamond or Daresbury. As with the Diamond rings, we use the centroid of the TTWA where Daresbury is located as origin (and similarly for the Newcastle-upon-Tyne rings discussed below).

research activity in relevant scientific disciplines would have been the same around Oxford and Daresbury in the absence of the construction of Diamond. In other words, any remaining time-varying unobservables, such as a general tendency for clusters to strengthen over time, are netted out by differencing the coefficients on D_{DI}^r and D_{DA}^r . Subsection 4.1 is careful to provide descriptive evidence in support of this identifying assumption.

In our context, two factors complicate the interpretation of the parameter estimates resulting from (5). The first complicating factor arises because, strictly speaking, (5) only provides estimates of the net treatment effect of distance to Diamond. If there are spillovers which raise research activity across the UK, e.g. because of global externalities (research advances), interactions (increased collaboration across UK universities) and general equilibrium effects (increase in the supply of researchers in the relevant fields), then we will underestimate the impact of distance to Diamond on the level of research activity. However, we still correctly capture the effect on the geographical distribution of activity, at least in a relative sense.

A second complicating factor works in the opposite direction. If the synchrotron at Daresbury had continued to operate, then comparing the coefficients on D_{DI}^r and D_{DA}^r would give us the impact of distance to Diamond. However, as we made clear above, the 2nd generation synchrotron at Daresbury was closed shortly after the opening of Diamond, so comparing the coefficients on D_{DI}^r and D_{DA}^r gives us the total effect of these two changes. In other words we might conflate the treatment effect of opening Diamond and the ‘distreatment’ effect of closing Daresbury.¹³ Assuming that these effects are opposite in sign, then estimating (5) will cause us to overestimate the treatment effect of Diamond. In practice this may not be a major problem because Diamond, as a 3rd generation Synchrotron, allowed for far more advanced research than the existing 2nd generation synchrotron at Daresbury. This implies that the location of the synchrotron did not simply move from Daresbury to Oxford, but a new type of facility was opened in Didcot that enabled researchers to conduct novel types of experiments which rendered the existing synchrotron technologically obsolete. Nevertheless, to test for this issue, we can use the existence of a third cluster of activity in Newcastle-upon-Tyne, based at the Institute for Cell and Molecular Biosciences. Using the same logic as before, this suggests that we estimate:

$$p_{at} = \alpha_a + \sum_t \delta_t D_t + \sum_r \gamma_{DI}^r [D_{DI}^r \times I(t \geq 2007)] + \sum_r \gamma_{DA}^r [D_{DA}^r \times I(t \geq 2007)] \\ + \sum_r \gamma_{NT}^r [D_{NT}^r \times I(t \geq 2007)] + \beta X_{at} + \epsilon_{at}, \quad (6)$$

where D_{NT}^r are a set of R ‘ring’ dummies which take value one if the area is within a given distance of Newcastle-upon-Tyne, zero otherwise and everything else is as before. Comparing the coefficients on D_{DI}^r to D_{NT}^r gives us the treatment effect of Diamond, while comparing the coefficients on D_{DA}^r to D_{NT}^r gives us the (dis)treatment effect of closing Daresbury. This identification relies on the assumption that on average unobservables vary with distance to the existing clusters (e.g. a general strengthening

¹³ We are grateful to Gabriel Ahfeldt for drawing our attention to this point.

of clusters) in a similar way across the three largest clusters of activity in the UK. We can check whether this is true for the third largest cluster by comparing our estimates from the diff-in-diff-in-diff specification (6) to the total effect estimated from the diff-in-diff specification that does not use the existence of a third cluster (5). For these reasons, (5) and (6) are our preferred specifications.

The use of Newcastle upon Tyne as an additional runner-up location was based on observable research output. As a final alternative to relying on Daresbury or another existing research cluster, we construct a synthetic control location using the approach proposed by Abadie and Gardeazabal (2003). That is, we construct an artificial location that resembles the area around Diamond in terms of its pre-treatment characteristics, including publication pre-trends, as closely as possible. This approach, therefore, accounts for any potential endogeneity in Diamond's location choice due to selection on pre-treatment observables.

The synthetic control location is a convex combination of all J available control locations. The combination of control locations depends on a $(J \times 1)$ vector of weights $\mathbf{W} = (w_1, \dots, w_J)$ with $0 \leq w_j \leq 1$ and $w_1 + \dots + w_J = 1$.¹⁴ The weights are chosen as follows. Let \mathbf{Z}_1 denote a $(K \times 1)$ vector of pre-treatment characteristics of the treated location including both observable characteristics and pre-treatment research outputs. Let \mathbf{Z}_0 denote a $(K \times J)$ matrix with the same variables for the J control locations.¹⁵ \mathbf{W} is chosen such that the pre-treatment characteristics of the synthetic control location $(\mathbf{Z}_0\mathbf{W})$ closely match the pre-treatment characteristics of the treated location (\mathbf{Z}_1) . Specifically, \mathbf{W}^* is chosen to minimise $(\mathbf{Z}_1 - \mathbf{Z}_0\mathbf{W})'\mathbf{V}(\mathbf{Z}_1 - \mathbf{Z}_0\mathbf{W})$ where \mathbf{V} denotes the relative importance of pre-treatment characteristics \mathbf{Z} . Using this framework, the effect of interest is simply computed as the difference between the actual and the synthetic counterfactual post-treatment outcomes $\mathbf{Y}_1 - (\mathbf{Y}_0\mathbf{W}^*)$. Note that the inclusion of pre-treatment outcomes in the set of pre-treatment characteristics means that we control for any differences in pre-treatment trends.

Unlike our main specifications, this synthetic control approach only allows us to estimate the treatment effect for a specific treatment group – in our case the area immediately around Diamond. In contrast to the intensity to treat approach used for our main results, this raises an additional inference problem given that there are only two locations – the treated and the control location. One way to address this is to carry out inference using placebo permutation tests (Abadie *et al.*, 2010, 2015). We estimate the publication output gap (the difference in output between the treated and synthetic control locations) iteratively for each of the other locations. That is, we proceed as if Diamond had been located in each location, estimate a corresponding synthetic

¹⁴ Requiring weights to sum to one avoids extrapolation outside of the support of the control locations.

¹⁵ We define pre-treatment characteristics to include research outputs for 2000 and 2006 as well as observable characteristics of locations for 2000–6: measures of local skill composition (%NVQ4 and above), the size of the local labour force, total grant payments from the STFC and both total grant payments and number of grants from the Wellcome Trust (i.e. the covariates X_{it} in (4)). In addition, we include a measure of research 'quality' and 'power' based on the 2001 Research Assessment Exercise (RAE) (see online Appendix B.3 for details). Since the RAE 2001 variables are based on research published between 1996 and 2001, accounting for the 2001 results provides us with a way of controlling for the information available to the government when making its location choice. Note that in (4), the effect of this time-invariant variable was absorbed in the fixed effects.

control, and compute the output gap. This results in a distribution of output gaps for all control locations. Comparing the output gaps between the location that was treated and all placebo locations provides a falsification test that shows whether the observed effect can be replicated by choosing any location at random.¹⁶

3. Data

The main data collection challenge consists in identifying relevant research input and output and its location. As explained above, we focus on scientific publications. Our starting point is a complete list of scientific publications that have resulted from work at Diamond. All users of the Diamond synchrotron are registered and report any scientific publication that results from allocated beamtime. The list contains 347 publications (as of December 2010) in 121 scientific journals. We refer to this set of publications as ‘Diamond articles’.

For these articles we collect the corresponding information on authors and their affiliations. We find that the 1,760 researchers listed as authors in these publications are affiliated to 441 institutions.¹⁷ Since author names and affiliations are not consistently formatted across the different journals, we standardised the data (as described in online Appendix B.1). Since we focus our analysis on publications by UK-based researchers, we drop articles that do not have at least one UK-affiliated author.¹⁸ This reduces the number of articles to 332 with 1,282 UK-based authors affiliated to 194 UK institutions. To determine the geographical location of researchers, we identify the postcodes of all UK affiliations and match it to the Code-Point data which contains National Grid co-ordinates.¹⁹

In a second step, we use the complete set of 347 Diamond articles to retrieve similar scientific publications with similarity defined by the overlap in cited references.²⁰ We pick the five most similar articles for each of our Diamond articles, yielding a total of 1,528 articles.²¹ We also collect related articles imposing the additional restriction that articles have to be published in either a field journal that pertains to the same field as

¹⁶ In our regression approach, we obtain the Diamond effect by comparing the effect around Diamond to all other locations with the effect around Diamond’s runner-up location to all other locations. Using the placebo permutation test, we obtain the effect by comparing the difference between the location around Diamond and the synthetic Diamond location to the differences between all other locations and their synthetic controls.

¹⁷ Different departments at the same university (e.g. the Departments of Physics and of Chemistry at the University of Oxford) are counted as different affiliations.

¹⁸ We also exclude Northern Irish affiliations to avoid complications in calculating appropriate distances for those authors.

¹⁹ Code-Point data is provided by Edina Digimap. The Code-Point data provides a precise geographical location for each postcode unit in the UK determined by its National Grid co-ordinates given by Easting and Northing values. Given the grid points for i and j , distances are calculated as $Distance = \sqrt{[Northing_i - Northing_j]^2 + [Easting_i - Easting_j]^2}$.

²⁰ Thomson Reuter’s Web of Science offers a search tool to identify such articles. Since we are interested in finding any scientific articles related to research conducted at Diamond, we do not restrict the base set of Diamond articles to those by UK-based authors.

²¹ Some articles are among the top five of several Diamond articles, which explains why this number is less than $347 \times 5 = 1,735$. We also experimented with alternative ways of retrieving related articles, for example based on keywords and abstracts. However, substantial differences across journals (e.g. only 54 out of the 121 journals report keywords), make these alternative procedures less suitable and they would require a greater amount of subjective assessment than our chosen method.

the original article (e.g. *Crystal Growth & Design* and *Acta Crystallographica* which are both crystallography field journals) or a general interest journal (e.g. *Science*). However, imposing this journal-based restriction means that articles are on average less similar to our Diamond articles in terms of reference overlap. We use this restricted set of similar articles to test the robustness of our results (see Section 5).

Since our definition of related articles rests on the similarity between Diamond and related articles, we also investigate concerns over the potential endogeneity of the location of Diamond with respect to the geographical distribution of those related articles. In particular, we consider the possibility that the set of related articles belongs to related fields of research that benefit directly from the synchrotron and that government accounted for this when deciding on the final location of Diamond. Results reported in Section 5 show that this form of endogeneity does not drive our results. To address this concern further, we also provide results for a third approach to the construction of related articles where we define all articles published in the same journal as a Diamond article as related. The definition makes no assumption on similarity between Diamond and related articles other than that articles are published by the same journal (see online Appendix B.2 for more details).

For all three sets of related articles, we then proceed as for Diamond articles: standardising author names and affiliations; keeping only articles with at least one UK-affiliated author; and identifying locations for these authors based on mapping postcodes using Code-Point data. We refer to the publications identified in this way as 'related articles' and their authors as 'related authors.' Online Appendix B.2 contains further details on identifying similar academic publications. Finally, note that in our analysis, we drop all authors employed directly at Diamond or Daresbury from the sample.²²

4. Results

In this Section, we consider results when using academic articles and their authors as our measure of research activity. We first offer some descriptive evidence and then discuss the main analytical results. The main results are based on TTWAs as the geographical unit of observation and use counts of academic articles as our measure for research output. Results for alternative specifications and output measures are reported in the robustness Section 5. We also discuss results on possible channels that drive the observed changes in scientific output at the extensive and intensive margins in Section 6.

4.1. Descriptives

As discussed in Section 3, we have 347 Diamond articles (based directly on experiments conducted at Diamond). There are 1,760 authors for these Diamond articles, with on

²² This avoids picking up any effect that comes from relocation of scientific staff from Daresbury to Oxford. Online Appendix C provides an analysis of any relocation effects. Note however that all our results are robust to including authors employed directly at the SRS or Diamond; including these authors in fact further strengthens the Diamond effect.

average, five authors per article. These authors are affiliated with 441 institutions worldwide with, on average, 4.2 affiliations per article. There are 1,282 authors with at least one UK affiliation (44% of all affiliations). As discussed in Section 2 – although we use all Diamond articles to identify related publications – we focus only on authors with British affiliations when considering the impact on the geographical distribution of research. This means dropping 15 articles with no British-affiliated authors, leaving us with 332 Diamond articles with 1,282 authors affiliated to 194 British institutions. Finally, after dropping all authors employed directly at Diamond or Daresbury from the sample, we are left with 311 Diamond articles.

Table 1 shows descriptive statistics for these 311 Diamond articles. There are, on average, 5.09 authors per article. Co-authors tend to share affiliations, so there are fewer institutional affiliations per article (the mean is 1.75). The median number of institutional affiliates per author is 1, although some authors have more (due to multiple affiliations or changes in institution). The Table shows data on the geographic distribution of researchers listed on Diamond articles in terms of distances (in kilometres) to Diamond and Daresbury pre and post-Diamond. These data suggest that, on average, authors are located considerably closer to Diamond pre-2007. Post-2007 distance to both Diamond and Daresbury is smaller than pre-2007: the average distance from Diamond is 172 kilometres pre-2007, 158 kilometres for 2007-on and

Table 1
Summary Statistics for Articles

	Mean	Median	SD	Minimum	Maximum
'Diamond articles'					
Descriptive statistics of authors & affiliations (UK only)					
No. of authors per article	5.09	5	2.72	1	16
No. of affiliations per article	1.75	1	0.95	1	6
No. of affiliations per author	1.14	1	0.39	1	3
Geographical distribution of 'Diamond' author affiliations					
<2007 (before establishment of Diamond)					
Distance (kilometres) to Diamond	171.7	107.9	144.8	1.2	521.4
Distance (kilometres) to Daresbury	226.3	250.6	68.9	7.1	363.3
≥2007 (after establishment of Diamond)					
Distance (kilometres) to Diamond	157.9	107.9	147.4	1.7	605.0
Distance (kilometres) to Daresbury	209.4	223.1	85.9	7.0	414.0
Related articles					
Descriptive statistics of authors & affiliations (UK only)					
No. of authors per article	3.83	3	2.40	1	17
No. of affiliations per article	1.62	1	0.88	1	7
No. of affiliations per author	1.26	1	0.50	1	4
Geographical distribution of author affiliations					
<2007 (before establishment of Diamond)					
Distance (kilometres) to Diamond	160.2	106.5	133.8	1.2	536.5
Distance (kilometres) to Daresbury	213.2	235.3	80.2	7.0	363.3
≥2007 (after establishment of Diamond)					
Distance (kilometres) to Diamond	156.2	107.9	151.3	1.7	605.8
Distance (kilometres) to Daresbury	214.1	224.7	84.9	7.0	415.2

Notes. There are 311 'Diamond articles,' 1,179 'Diamond authors', that are affiliated to 194 institutions. There are 510 related articles, 1,235 'related authors', that are affiliated to 222 institutions.

the average distance from Daresbury is 226 kilometres pre-2007 and 209 kilometres post-2007.

The lower panel of Table 1 shows descriptive statistics for related articles. On average, there are 3.83 authors per article, with co-authoring again favouring same institution, so that the number of affiliations per article is 1.62. The mean and median distances for these authors are both close to that of Diamond authors, with authors being located closer to Diamond post-opening (on average 4 kilometres closer). Average distances to Diamond are in fact slightly smaller pre-Diamond than for Diamond authors.

The map in Figure 2 shows the location of Diamond and related publications pre- and post-Diamond opening. The map plots the location of Diamond and Daresbury; the distance rings; and the number of publications (counted only once regardless of the number of authors) in a given location (as determined by the affiliation postcode) summed over two periods: pre-Diamond (2003–6) and post-Diamond (2007–10). We see some research activity around both Daresbury and Diamond pre-Diamond. Comparison between 2003–6 and 2007–10 shows clear evidence for increased activity around Diamond and much less so around Daresbury. Note that this disproportionate increase in research activity around Diamond appears to be particularly pronounced in direct proximity of Diamond (and particularly within the 25 kilometres distance band).

Figure 3 offers additional preliminary evidence for a proximity effect post-Diamond. The Figure shows the number of related articles by authors located within 25 kilometres of Diamond (solid line) or Daresbury (dashed dark grey line).²³ The Figure highlights that the number of academic articles published by researchers close to Diamond increases significantly shortly after Diamond was opened, whereas the increase is a lot less steep around Daresbury. This evidence strongly suggests a strong proximity effect within 25 kilometres of Diamond. Because Figure 3 uses only data for related publications, the plot provides strong, descriptive evidence for local externalities created by Diamond.

To explore this further, Figure 4 shows annual coefficient estimates (γ_{Ct}) from the regressions $p_{at} = \alpha_a + \sum_t \delta_t D_t + \gamma_{Ct} [D_C^{25} \times D_t] + \beta X_{at} + \epsilon_{at}$ (with $C = [DI, DA]$) for Diamond and Daresbury (within 25 kilometres) where $t = 2000, 2001, \dots, 2010$ and 2007 (the year of Diamond's opening) is the omitted category. These regressions provide direct evidence on our identifying assumption of comparable pre-Diamond trends in both locations (regardless, we include pre-trends in all regressions). Whereas there is only a very moderate effect on number of publications after 2007 in the Daresbury area, the Diamond Figure shows significant increases in academic activity within 25 kilometres of Diamond. The remainder of this Section estimates the specifications discussed above to make these comparisons more precise.

4.2. Main Results

We start by providing estimation results for (3) and (4) in columns (1) and (2) of Table 2. The dependent variable is the TTWA-by-year count of scientific articles

²³ In the Figures, we focus on related articles to allow us to consider trends before the opening of Diamond.

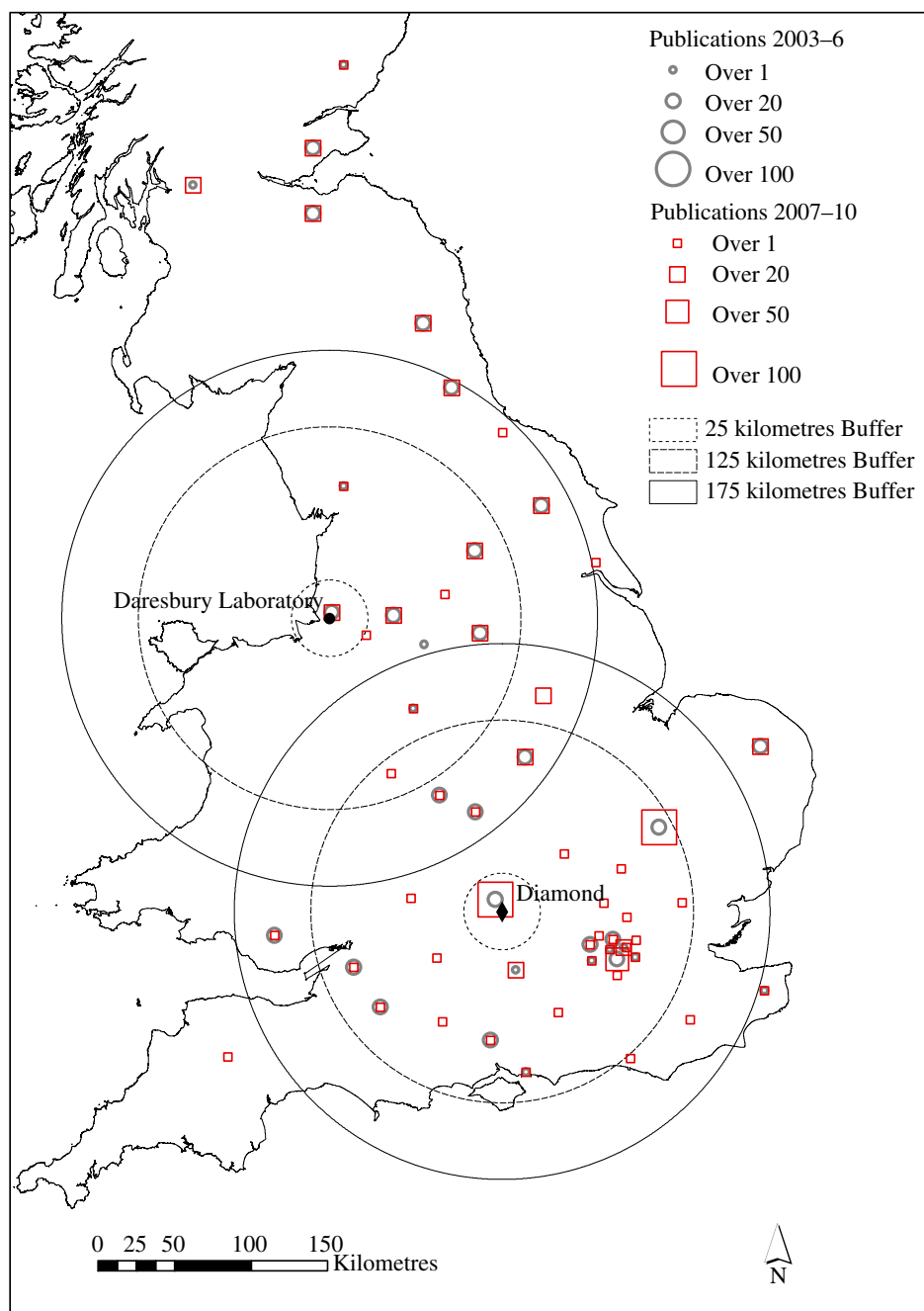


Fig. 2. *Author Counts for Diamond and Related Articles – Pre-Diamond (2003–6) versus Post-Diamond (2007–10)*

Notes. The map plots the number of unique publications in a given location (as determined by the postcode of the authors' affiliation) before the opening of Diamond (summed over the period 2003–6) as well as after the opening of Diamond (summed over the period 2007–10).

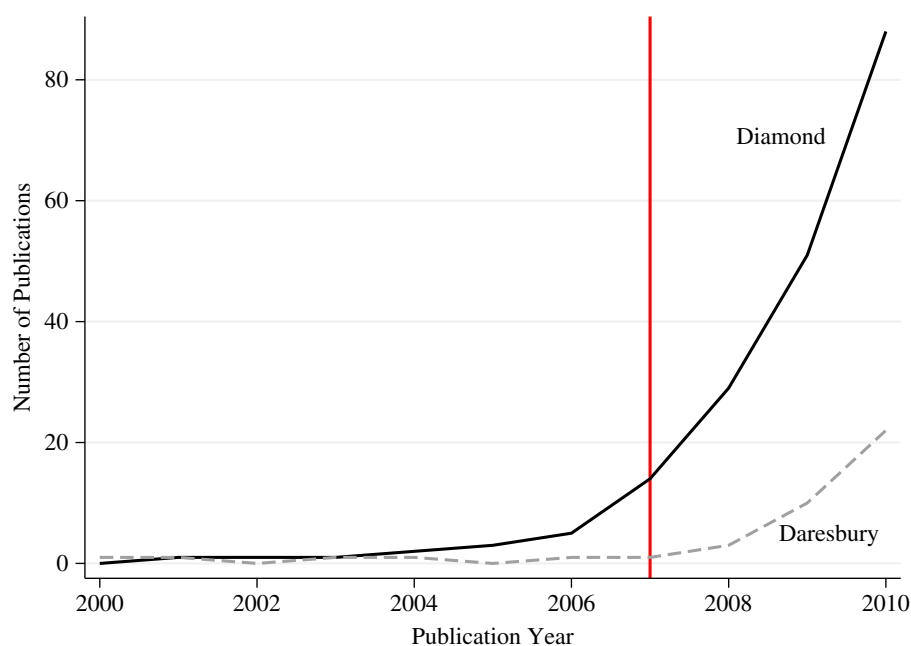


Fig. 3. *Number of Related Articles Within 25 Kilometres Distance Ring of Diamond and Daresbury (Before After 2007)*

(i.e. ignoring the number of authors in a given location) for the period 2000–10.²⁴ Column (1) reports OLS fixed effects results when we include time dummies, the three ring dummies corresponding to distances 0–25, 25–125, 125–175 kilometres (with the omitted category more than 175 kilometres) interacted with a post-Diamond dummy – an indicator for years after the opening of Diamond. The interaction of the ring dummies with the post-Diamond dummy captures the way that the impact of Diamond varies with distance to Diamond. As is clear from column (1) there is a strong positive effect at short distances – the coefficient on the interaction term for the 0–25 kilometres ring is large and statistically significant (at the 1% level). There is no effect at 25–125 or 125–175 kilometres.

As discussed above, one possible explanation of these results is that locations differ in terms of their research potential and that this research potential may explain both the location of Diamond and any differences in research activity. The specification in column (2) of Table 2 deals with this possibility by introducing our set of observable covariates to capture differences in TTWA research potential. As a result of introducing these variables, the coefficient on the first distance ring around Diamond drops in magnitude but is still statistically significant at 1%.

In columns (3) and (4), we show results when we also introduce a set of ring dummies to control for distance to Daresbury – the runner up location (5). Comparing coefficients on the two sets of ring dummies gives an estimate of the distance effect that

²⁴ Online Appendix Table E2 contains the corresponding descriptive statistics.

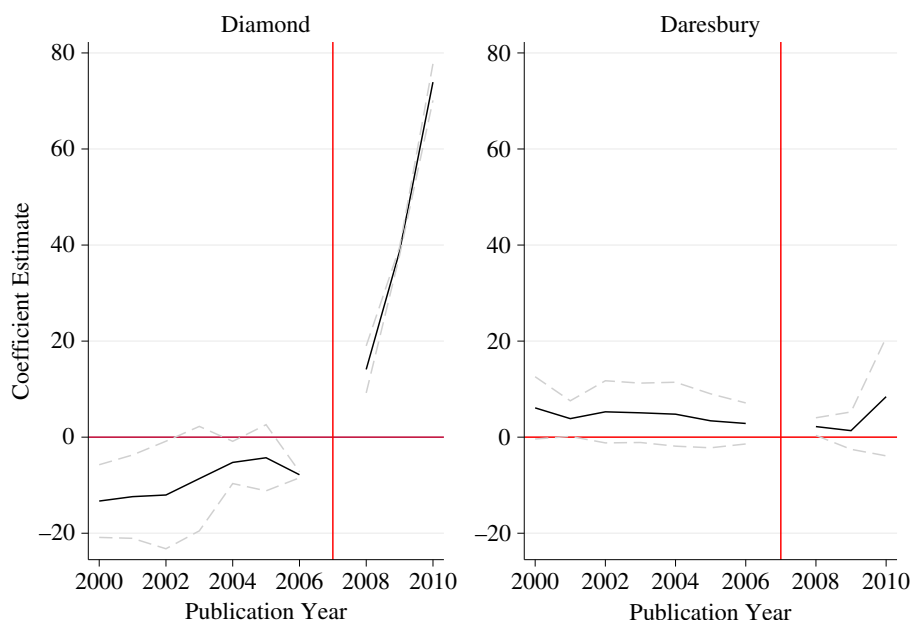


Fig. 4. Number of Related Articles – Annual Coefficient Estimates for Diamond and Daresbury (Before & After 2007)

Notes. The two graphs show annual coefficient estimates γ_{Ct} from the regressions $p_{at} = \alpha_a + \sum_t \delta_t D_t + \gamma_{Ct} [D_C^{25} \times D_t] + \beta X_{at} + \epsilon_{at}$ (with $C = [DI, DA]$) for Diamond and Daresbury (within 25 kilometres) where $t = 2000, 2001, \dots, 2010$; 2007 is the omitted category.

(partially) controls for unobservables that may be correlated with distance to Diamond as would happen, for example, if the tendency of existing clusters to specialise was strengthening over time and that this tendency drove the Diamond location decision. In column (3), we see that the strong positive effects close to Diamond after 2007 are not replicated around Daresbury, although we find a small positive and statistically significant coefficient on the first Daresbury distance ring. However, column (4) shows that this finding is not robust to the introduction of observable time varying characteristics of locations that capture research potential and research funding allocations. The coefficient on the first Daresbury distance ring is now negative and no longer statistically significant. If the positive effect of Diamond is driven purely by the tendency for existing clusters to strengthen over time, then we should observe a similar pattern of increased activity in areas close to the centre of the alternative cluster in Daresbury – and these results suggest that we do not. Instead, we find a positive, statistically significant difference between the 25 kilometres ring dummies around Diamond and Daresbury.

At the bottom of Table 2, we also report confidence intervals estimated using the Conley and Taber (2011) approach. Conley and Taber (2011) point out that classical inference can be misleading in the difference-in-differences estimation when the number of treated units is small. They propose an alternative simulation-based method of inference that constructs standard errors and confidence intervals using the estimated residuals of the control groups. Our findings are robust to this alternative

Table 2

The Effect of Distance to Diamond versus Distance to Daresbury (209 TTWAs; 2000–10)

	(1)	(2)	(3)	(4)
Diamond $\times I(t \geq 2007)$				
25 kilometres	19.777*** (0.059)	13.892*** (0.802)	19.879*** (0.042)	13.971*** (0.724)
125 kilometres	1.417 (0.884)	−0.419 (0.451)	1.519* (0.884)	−0.369 (0.457)
175 kilometres	−0.0005 (0.528)	−0.707 (1.026)	0.101 (0.527)	−0.659 (1.026)
Daresbury $\times I(t \geq 2007)$				
25 kilometres			1.546*** (0.322)	−0.879 (1.598)
125 kilometres			1.699* (0.723)	1.350* (0.798)
175 kilometres			0.933 (0.835)	0.143 (0.539)
Time dummies	Yes	Yes	Yes	Yes
Linear pre-trend	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes
Covariates	No	Yes	No	Yes
R ²	0.308	0.279	0.308	0.278
Observations	2,299	2,299	2,299	2,299
$\theta^{r=1} = \gamma_{DI}^{r=1} - \gamma_{DA}^{r=1}$			18.333*** (0.319)	14.850*** (1.984)
Conley-Taber 95% confidence intervals				
Diamond $\times I(t \geq 2007)$	[12.23; 20.80]	[6.37; 14.94]	[12.33; 20.90]	[6.45; 15.02]
25 kilometres				
Daresbury $\times I(t \geq 2007)$			[−8.19; 2.65]	[−10.59; 0.25]
25 kilometres				

Notes. Dependent variable is publication count by TTWA and year. Robust standard errors clustered at TTWA level. All regressions include a constant. Covariates include %NVQ4 and above, Labour force, No. of Wellcome grants, £'000 awarded by Wellcome, £'000 awarded by STFC. Conley and Taber (2011) inference approach to account for small number of treated TTWAs within 25 kilometres distance rings. * significant at 10%, ** at 5%, *** at 1%.

method of inference; the confidence intervals reported at the bottom of Table 2 show that the coefficient on the first distance ring around Diamond is still highly statistically significant whereas the effect around Daresbury is indistinguishable from zero.

In terms of magnitudes, given that we are using a simple article count, ignoring the number of authors in a given location, the point estimate of 13.971 in column (4) is equivalent to roughly 8.36 extra articles per year produced within 25 kilometres of Diamond.²⁵ We see around 75 Diamond and related articles published per year (821 articles over an 11 year period) so this corresponds to an increase of a little over 11% of the total within 25 kilometres of Diamond.

Recall, however, that Daresbury may be affected by a negative distreatment effect that occurs from the shut-down of the second generation synchrotron that used to

²⁵ We calculate this Figure using the average number of affiliations per article (1.67) ignoring the fact that some authors report multiple affiliations – some of which may be in the same location. Given that the average number of affiliations per author is around 1.18 this should not represent a major source of bias.

operate on that site. If we view these as related decisions and we are interested in the net effect, then the estimates so far still provide the coefficients of interest. If we are interested in the gross impact then this remains a concern – even though there is reason to be somewhat sceptical of this possibility given the big differences between the 2nd and 3rd generation technologies. To address this issue, we estimate (6) which includes an additional set of Newcastle ring dummies to provide an alternative set of distance coefficients to compare to those for Diamond – this time these comparison ring dummies capture post-Diamond changes in unobservable characteristics which are correlated with distance to the Institute for Cell and Molecular Biosciences in Newcastle (a third cluster of related research in the UK). Column (1) in Table 3 reports results (for the interacted terms) when including the Diamond and Newcastle ring dummies. The coefficients on the Newcastle dummies are positive and statistically significant. Note, however, that if we use the Conley and Taber (2011) method of inference, results reported in the last row of the Table show that the effect around Newcastle is not statistically significantly different from zero. Column (2) shows that these results do not change when we include all three sets of ring dummies, while column (3) allows a direct comparison of Daresbury and Newcastle and shows tentative evidence of a distreatment effect for Daresbury. Overall, these results mean that, even accounting for a general tendency of clusters (in fields relevant to Diamond) to strengthen over time, we find a strong positive effect of Diamond on research output in the area close to Diamond.

As an alternative to Daresbury and Newcastle, we use the pattern of results so far (which suggest a strong effect within 25 kilometres of Diamond) to justify an alternative comparison of the effect of Diamond based on a synthetic control location. Details of the methodology are provided in subsection 2.2. Online Appendix Table E1 compares the characteristics of the areas around (within 25 kilometres) Diamond and Daresbury with those of the synthetic control area.²⁶ The Table shows that Daresbury resembles the Diamond location particularly well in terms of research quality, measured by the grades awarded to departments by the RAE. Daresbury fares less well with regard to scale, that is, in terms of the total number of grants and total value of grants awarded by the Wellcome trust as well as the number of world-class researchers (as assessed by the RAE). In this regard, the synthetic combination of different locations resembles the Diamond location a lot better. Figure 5 shows the comparison of the publication output by researchers in direct proximity to Diamond and output constructed for the synthetic control location. By construction, the synthetic control location matches the pre-treatment publication trend closely but publication counts diverge substantially once Diamond becomes operational. The plot on the right-hand-side shows the publication gap between Diamond and the synthetic control. The gap is almost zero throughout the 2000–6 period but jumps immediately once Diamond opens and keeps increasing rapidly. This offers strong evidence in favour of our previous findings based on the Diamond-Daresbury comparison. Figure 5 also compares the publication output of the synthetic control area directly with the Daresbury area. The Figure shows that publication counts behave similarly before

²⁶ The synthetic control location is constructed based only on Cambridge (0.368), Edinburgh (0.351), London (0.156) and Newcastle (0.125) with corresponding weights in brackets.

Table 3

The Effect of Distance to Diamond versus Distance to Daresbury versus Distance to Newcastle (209 TTWAs; 2000–10)

	(1)	(2)	(3)
Diamond $\times I(t \geq 2007)$			
25 kilometres	14.088*** (0.769)	14.123*** (0.706)	
125 kilometres	-0.301 (0.430)	-0.317 (0.455)	
175 kilometres	0.210 (0.789)	-0.618 (1.023)	
Daresbury $\times I(t \geq 2007)$			
25 kilometres		-0.839 (1.594)	-1.090 (1.690)
125 kilometres		1.390* (0.798)	1.268 (0.773)
175 kilometres		1.778*** (0.230)	-1.306 (1.129)
Newcastle $\times I(t \geq 2007)$			
25 kilometres	3.517*** (0.545)	3.538*** (0.525)	3.283*** (0.772)
125 kilometres	-0.475 (0.428)	-0.460 (0.434)	-0.519 (0.434)
175 kilometres	0.977 (0.956)	0.784 (1.307)	0.600 (1.412)
Time dummies	Yes	Yes	Yes
Linear pre-trend	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
R ²	0.277	0.276	0.286
Observations	2,299	2,299	2,299
$\theta_1^{r=1} = \gamma_{DI}^{r=1} - \gamma_{DA}^{r=1}$		14.963*** (1.944)	
$\theta_2^{r=1} = \gamma_{DI}^{r=1} - \gamma_{NT}^{r=1}$	10.571*** (0.504)	10.585*** (0.493)	
$\theta_3^{r=1} = \gamma_{DA}^{r=1} - \gamma_{NT}^{r=1}$		-4.377** (1.919)	-4.374** (2.077)
Conley–Taber 95% confidence intervals			
Diamond $\times I(t \geq 2007)$	[6.57; 15.14]	[6.60; 15.17]	
25 kilometres			
Daresbury $\times I(t \geq 2007)$		[-10.55; 0.29]	[-8.39; 0.08]
25 kilometres			
Newcastle $\times I(t \geq 2007)$	[-3.95; 4.62]	[-3.93; 4.64]	[-3.93; 4.49]
25 kilometres			

Notes. Dependent variable is publication count by TTWA and year. Robust standard errors clustered at TTWA level. All regressions include a constant. Covariates include %NVQ4 and above, Labour force, No. of Wellcome grants, £'000 awarded by Wellcome, and £'000 awarded by STFC. Conley and Taber (2011) inference approach to account for small number of treated TTWAs within 25 kilometres distance rings. *Significant at 10%, **at 5%, ***at 1%.

Diamond opens; in 2007, when Diamond opens, the synthetic control matches the sudden increases in publications more closely than Daresbury. Daresbury, in contrast matches the eventual strong increase in publications better than the synthetic control, which sees a flattening of publications instead of a continued rise. Finally, Figure 6 shows the results from the placebo test described in subsection 2.2 above. The

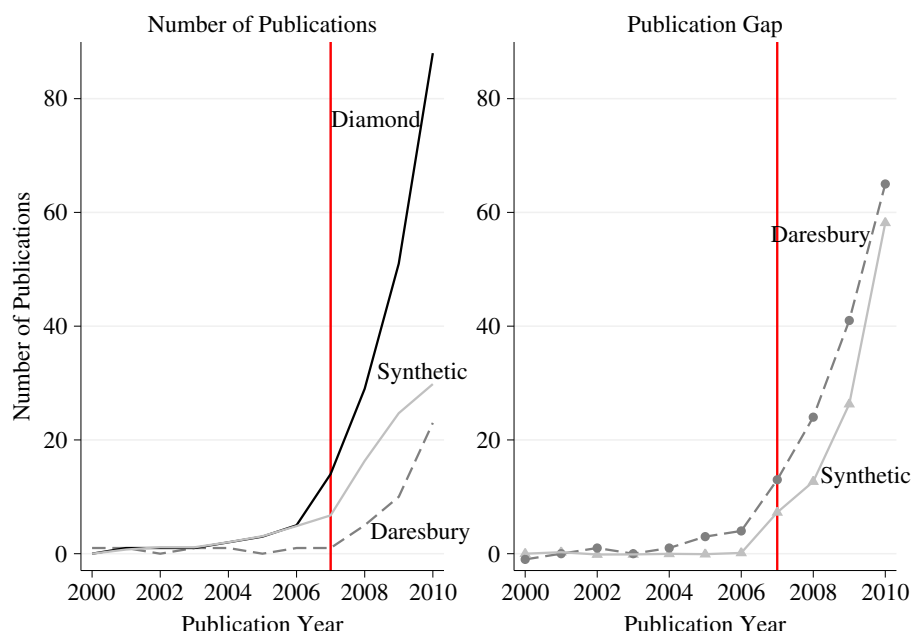


Fig. 5. *Diamond versus Daresbury versus Synthetic Control Location*

Notes. The left-hand-side graph shows publication counts for the 25 kilometres distance ring around Diamond and Daresbury as well as the synthetic control location. The right-hand-side graph shows the difference in publication counts between Diamond and Daresbury as well as the difference in publication counts between Diamond and the synthetic control location. The synthetic control location matches the pre-Diamond scientific publication output based on the following variables: %NVQ4 and above, Labour force, no. Wellcome grants, £'000 awarded by Wellcome, £'000 awarded by STFC, no. 5/5* researchers Biological Sciences, Grade Biological Sciences, no. 5/5* researchers Chemistry, Grade Chemistry, no. 5/5* researchers Physics, Grade Physics and the no. of publications in 2000 and 2006.

Figure plots the output gap between Diamond and its synthetic control (black line) as well as that of every other location and its corresponding synthetic control (grey lines). The Figure, therefore, allows us to gauge whether the effect observed for the area around Diamond is similar to that of any of the other locations. Figure 6 strongly rejects this, the estimated gap for Diamond clearly stands out from all other estimates.

The synthetic control method also allows us to compare the Daresbury and the synthetic counterfactual outcomes. We compare directly the number of publications for the synthetic control location to the publications produced by researchers in the Daresbury location (i.e. within 25 kilometres). The synthetic control location cannot do worse than Daresbury in replicating the pre-treatment publication output (the synthetic control location can always replicate Daresbury by assigning a weight equal to one to the Daresbury location). Therefore, differences in the publication gap once Diamond opens between Diamond and Daresbury and Diamond and its synthetic control location reflect on the one hand Daresbury's (lack of) similarity to Diamond in terms of observable pre-treatment characteristics and on the other the presence of unobservable time-varying characteristics not taken into account by the synthetic control location.

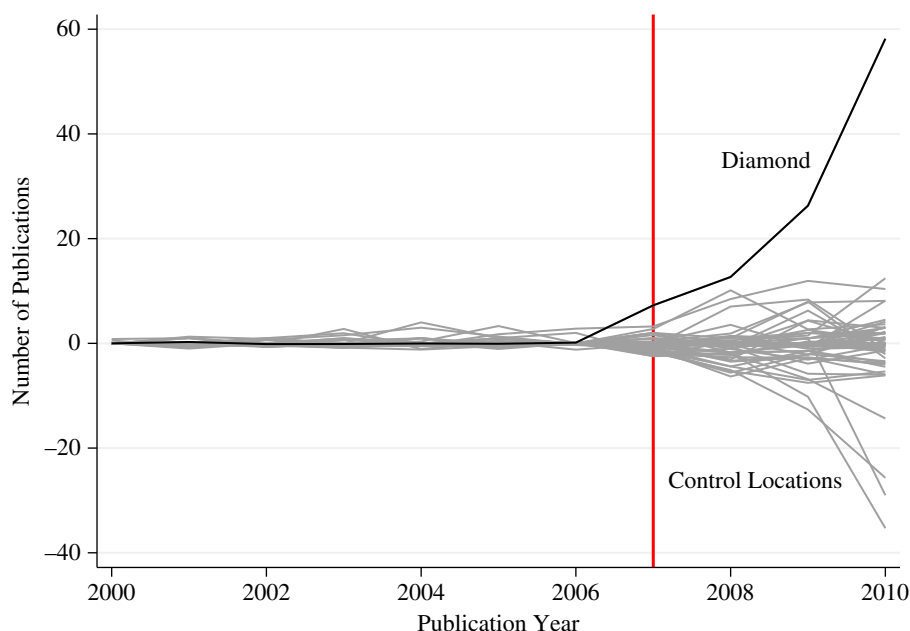


Fig. 6. *Gap Between Treated Location and Synthetic Control – Diamond versus Placebo Treatments*
Notes. The graph shows the gap between the number of publications of a treated location and its synthetic control. The gap for Diamond is obtained as in Figure 5. The gaps for all other locations are obtained by iteratively assigning treatment to each of the other locations and estimating its synthetic control.

5. Robustness

In this Section, we report results from a number of robustness checks. First, we use different specifications for the distance rings (both in terms of width and number of rings). Second, we consider a number of different ways of weighting and constructing article counts. Third, we vary the way in which we construct the sample of Diamond related research. Finally, we show results for alternative units of observation. All of these variations leave our main results essentially unchanged.

Online Appendix Tables E4 and E5 show the extent to which our results are robust to changing the width of the distance rings and the number of distance rings. The impact of Diamond is highly localised, it is confined to the area between Oxford University and Diamond, which corresponds to the 25 kilometres distance ring. However, the specification of the remaining rings is not so important. When we switch to only two rings (0–25, 25–125 kilometres) we continue to find positive significant effects within 25 kilometres but not beyond that. Switching to four rings (0–25, 25–75, 75–125, 125–175 kilometres) shows the same overall pattern – positive effects within 25 kilometres, very much smaller or zero effects further away.

Our main results rely on a simple count of articles ignoring the number of authors in a given location (i.e. if two or more authors of an article are located in Cambridge, the article is only counted once for Cambridge). Table E6 reports results using a number of alternative ways of counting scientific publications to create the TTWA output

variable. Column (1) reports results when we count each article once for each author (i.e. if two authors of an article are located in Cambridge, the article is counted twice for Cambridge). Column (2) reports results when we weight publications by the number of authors where each author is given the same weight equal to one divided by the number of authors. In column (3), we instead allocate each author a weight that corresponds to their relative importance for a given publication. In the sciences, by convention the position of an author in the list of authors is a strong indicator of the author's role in the research that has led to the publication.²⁷ This means by taking into consideration the ordering of authors in a given publication, we are able to weight author counts by authors' relative importance for a given research output. Column (4) instead uses a simple count of articles for a given location as many times as authors at distinct institutions are located in that TTWA (i.e. if two authors of an article are located in different institutions in Cambridge, we count the same article twice in Cambridge). Column (5) counts only articles by the first author.²⁸ Finally, column (6) counts articles dropping any authors that report multiple affiliations (avoiding concerns about correctly identifying an author's primary affiliation). As is clear from the Table, the overall pattern of coefficients on the distance rings is unchanged (although the magnitudes of the point estimates change as the total number of article counts changes across the different specifications) – for the corresponding descriptive statistics see online Appendix Table E2.

Next we investigate the potential endogeneity of our related articles. Specifically, the concern is that if our related articles belong to fields of research that benefit from a synchrotron and if the area around Oxford was particularly promising in these research fields before Diamond opened, our definition of related articles would bias us towards finding a localised Diamond effect. Furthermore, if this strength in specific fields around Oxford influenced the government's and Wellcome's decision to place Diamond in Didcot, one would expect to observe more articles in these fields written by authors located in the Oxford area after Diamond opened. In practice, two sets of results reported in Table E7 show no evidence for the endogeneity of our set of related articles suggesting that this does not drive our results.

For the first set of results, we allocate all Diamond and related articles into research fields (as defined by Thomson Reuters based on the journals in which the articles were published). We then ask whether the likelihood that a related article is in the same field as its corresponding Diamond article is a function of the related article's distance (as defined by its authors' affiliations) to Diamond. The results are shown in Table E7, columns (1)–(4). There is no evidence that proximity to Diamond of the related articles' authors makes it more likely that the related article will be in the same field as the corresponding Diamond article. In other words, there is no evidence to suggest that related articles 'located' close to Diamond are more likely to belong to the same

²⁷ The first and last author are usually perceived to have contributed the most to an article whereas authors appearing in the middle receive less credit (for survey evidence on this perception see Wren *et al.*, 2007). This means that we assign the first and last authors the same score whereas the score drops the further down an author name appears in the byline; i.e. we divide each author's share of a publication by the author's position in the byline, where the last author is treated in the same way as the first author.

²⁸ This requires us to discard a number of publications where the first author is not in our sample.

field as the corresponding Diamond article (as we would expect to find if the endogeneity problem applied).

For the second set of results, we randomly draw three ‘control articles’ for each related article. These control articles are drawn from the same journal as the related article and published in the same year. We then ask whether the likelihood that a given article is classified as a related article (as opposed to a ‘control related’ article) is a function of the article’s distance to Diamond. If our set of related articles was endogenous, proximity to Diamond should predict whether a given article is in fact a related article. The results reported in Table E7 columns (5)–(6) do not provide any evidence in favour of this hypothesis – proximity to Diamond does not predict whether an article is a related article. Columns (7)–(10) show that this is true regardless of whether a related article was published in a ‘Diamond journal’ or any different journal.

In addition to investigating the potential endogeneity of our related articles, we also adopt two alternative approaches to constructing the set of related articles. As explained in Section 3, our main results are based on a set of similar articles where similarity is defined by the overlap in cited references with Diamond articles. For columns (1)–(3) in Table E8, we construct the set of related articles while restricting the set of articles to those that have been published in either a field journal that pertains to the same field as the Diamond article or a general interest journal. This restriction, however, means that the set of related articles will be on average less similar in terms of the reference overlap. Collecting related articles in this way produces a set of 539 related academic articles with 1,519 authors that are affiliated to 231 institutions. The results are very similar to those displayed in Tables 2 and 3, that is, the coefficients on the first Diamond ring dummy are positive and statistically significant. Columns (4)–(6) in Table E8 show results when we construct the set of related articles avoiding our similarity requirement altogether and define any other article published in a journal in which a Diamond article has been published as potentially related. That is, we pick the top-3 Diamond journals, extract all articles with at least one UK-based author between 2000 and 2010 – resulting in a total 6,189 articles (see online Appendix B.2) – and estimate our main specifications with this alternative set of articles. The results shown in Table E8 are remarkably similar to those shown in Tables 2 and 3.

For our final set of robustness checks, we change the unit of observation – moving down to smaller local authorities (LAs) and even further down to institutions. Results are reported in Table E9. Columns (1)–(3) report results when using LAs. We find a large, positive effect within 25 kilometres of Diamond. The difference $\gamma_{DI}^{r=1} - \gamma_{DA}^{r=1}$ is positive and statistically highly significant. Columns (4)–(6) report results when we use universities/research institutions as our unit of analysis. Again the results for the 25 kilometres distance ring around Diamond remain qualitatively unchanged compared to the LA and TTWA-level analysis.

6. Mechanisms

Having established that our main finding – a strong positive effect of Diamond on institutions (and hence TTWAs) located within 25 kilometres of Diamond – is robust to a range of variations in specification, sample construction and units of observation, we now turn to potential drivers of this effect.

A useful starting point is to distinguish between direct and indirect effects of Diamond as discussed above. The framework presented in subsection 2.1 suggests a straightforward way to separate these two effects. As discussed in Section 3, we have data on articles that resulted directly from work at Diamond and related articles (where related means that scientists did not use Diamond for their research). Defining p_{at} as a combination of both types of article allows us to obtain an estimate of the combined effects of Diamond, that is, the combined effect of S_{at} and R_{at} as defined in (1). However, if we limit p_{at} to related articles, the coefficients on the ring dummies measure only the indirect effect R_{at} because this research did not use the synchrotron directly (and similarly in terms of the direct effect, if we limit p_{at} only to Diamond articles). Hence, we can obtain estimates of both the direct and indirect effects depending on how we define p_{at} .

Table 4 presents results when we distinguish between direct and related research output. For the direct effect, by definition, there are no articles published before the

Table 4
Direct versus Indirect Effect (38 and 209 TTWAs; 2000–10)

	Diamond only			Related only		
	No. of articles	Productivity	No. of authors	No. of articles	Productivity	No. of authors
	(1)	(2)	(3)	(4)	(5)	(6)
Diamond $\times I(t \geq 2007)$						
25 kilometres	42.742*** (0.900)	0.249*** (0.035)	106.206*** (2.062)	14.443*** (0.716)	0.026 (0.049)	33.151*** (1.337)
125 kilometres	5.960 (4.425)	0.384*** (0.079)	17.031 (13.091)	-0.069 (0.351)	0.002 (0.052)	-0.808 (1.093)
175 kilometres	-2.001 (1.687)	0.180** (0.077)	-3.869 (3.721)	0.051 (0.869)	0.463 (0.269)	-1.786 (2.419)
Daresbury $\times I(t \geq 2007)$						
25 kilometres	4.156 (5.053)	0.722* (0.421)	9.981 (13.260)	-0.243 (0.925)	-0.104 (0.087)	-0.034 (2.230)
125 kilometres	4.659** (1.844)	0.289*** (0.091)	13.168** (5.607)	1.821 (1.275)	0.008 (0.089)	3.589 (2.414)
175 kilometres	1.911 (2.215)	0.345*** (0.077)	3.990 (5.180)	0.512 (0.484)	0.287 (0.057)	0.621 (1.380)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Linear pre-trend	No	No	No	Yes	Yes	Yes
Covariates	No	No	No	Yes	Yes	Yes
Fixed effects	No	No	No	Yes	Yes	Yes
R ²	0.447	0.463	0.389	0.298	0.053	0.303
Observations	168	168	168	2,299	2,299	2,299
$\theta^{r=1} = \gamma_{DI}^{r=1} - \gamma_{DA}^{r=1}$	38.586*** (4.921)	-0.472 (0.416)	96.225*** (13.022)	14.687*** (1.318)	0.130 (0.104)	33.186*** (2.821)

Notes. Dependent variable in columns (1) and (4) is publication count by TTWA and year, in columns (2) and (5) the publication count divided by author count by TTWA and year, and in columns (3) and (6) unique author count by TTWA and year. Robust standard errors clustered at TTWA level. All regressions include a constant. Covariates include %NVQ4 and above, Labour force, No. Wellcome grants, £'000 awarded by Wellcome, and £'000 awarded by STFC. * Significant at 10%, ** at 5%, *** at 1%.

opening of Diamond. To allow for this, column (1) reports results when we estimate our basic specification using data on Diamond publications from 2007 onwards. As the Table shows, the number of observations drops because of the shorter time period. In addition, it no longer makes sense to use the ring dummies interacted with the post-Diamond dummy as all data is post-Diamond (although for ease of comparison, we report results as if we had imposed the redundant interaction). For Diamond articles, the results show a positive significant effect for the 25 kilometres ring (as for all articles). Column (4) presents results for the related articles using data for the whole time period (2000–10). As for total articles, the coefficient on the 25 kilometres Diamond ring is positive and statistically significant (and statistically significantly different from the coefficient on the 25 kilometres Daresbury ring). This suggests that Diamond affects not only research that relies on access to the facility directly, but also related research and hence provides strong evidence for the presence of local externalities created by Diamond.

In broad terms, the alternative explanations for the effects of Diamond can be split into two groups – either they result from an increase in inputs or from an increase in productivity (or some combination of both). We consider several such input and productivity effects as the rest of this Section explains. Our starting point are results reported in columns (2) and (5) of Table 4 which look for evidence of a productivity effect by running our main specifications using TTWA-by-year articles per author as the dependent variable. We distinguish between Diamond and related articles (with the restriction that this implies for the Diamond article specification, as discussed above). The results, in column (2) show a positive effect within 25 kilometres for Diamond articles but the coefficient is not statistically significantly different from that around Daresbury. For related articles (column (5)), we find no evidence of a productivity effect.

This suggests that our effects must be predominantly explained by changes in inputs. To consider this mechanism, results in columns (3) and (4) of Table 4 show what happens when we ignore the number of publications by author and instead use as our dependent variable author counts by TTWA and year (i.e. we count authors only once, independently of their number of publications in a given year). We view this as a measure of research input rather than output, which allows us to ask whether the Diamond effect can be in part explained by an increase in the number of scientists in proximity to Diamond (i.e. change at the extensive margin). Once again we distinguish between Diamond and related articles. Table 4 shows that we find a strong, positive, and statistically significant effect on the number of researchers within the 25 kilometres Diamond distance ring for both Diamond (column (3)) and related articles (column (6)). We find no such effect for Daresbury and a test for equality of the Diamond and Daresbury 25 kilometres ring coefficients is rejected at the 1% level. These results imply that one part of the explanation of our main results is the increase in the number of scientists working on relevant research in proximity to Diamond.

Next we ask whether this increase in the number of scientists can be partly explained by researchers that move closer to Diamond when the synchrotron facility became operational. We identify authors of related articles that move through changes in their affiliations provided by articles and only consider scientists that move between

geographically distinct institutions after Diamond was opened.²⁹ Figure D1 (in the online Appendix) shows the change in distances to Diamond and to Daresbury for scientists that move with the numbers denoting individual scientists.³⁰ In the left-hand panel, a given scientist is located above the 45° degree line if they moved closer to Diamond after Diamond opened and *vice versa*. The right-hand panel shows the same plot for distance to Daresbury and suggests that on average scientists moved away from Daresbury post-Diamond (for these scientists the average distance increases from 179 to 211 kilometres post-Diamond). The pattern is less conclusive for Diamond as shown in the left-hand panel. While 16 scientists moved closer to Diamond, 12 scientists moved away – although the effect of the former outweighs the latter with the average distance for these scientists decreasing from 168 to 134 kilometres post-Diamond. Focusing more specifically within a 25 kilometres radius of Diamond we see five scientists moving to be within that distance, and exactly the same number moving further away. In short, the number of movers is small (accounting for 4.4% of all ‘related authors’ in our sample and 8.2% of ‘related articles’) and movers generate zero net changes within 25 kilometres of Diamond so the indirect effect of Diamond is not explained by the relocation of scientists within the UK post-opening.³¹

However, this only considers domestic relocations, rather than the movement of scientists from abroad. To consider this, we look at authors of Diamond or related articles that change affiliation from abroad to the UK during our time period. Figure D2 (in the online Appendix) looks at the percentage of these authors that locate within 25 kilometres of either Diamond or Daresbury (distinguishing between Diamond authors and related authors). For ‘Diamond authors’ we see an increase in the share of authors from abroad that move to within 25 kilometres of Diamond around the time Diamond opens, whereas there is no such increase within 25 kilometres of Daresbury. That said, we also see an increase in the share of international movers already in a single year 2004, for which there is no apparent explanation. The pattern is very similar for authors of related articles. Overall, these results suggest that the increase in scientists around Diamond was partially driven by the location decision of scientists moving to the UK from abroad, although the magnitude of the effect only explains a fraction of the overall effect.

These results for movers suggest that the direct and indirect effects at the extensive margin are predominantly driven by ‘new’ scientists. The data on Diamond and related authors is of limited use in considering the mechanisms behind these effects because they only cover scientists that have used Diamond or published a related article. To consider the role of proximity to Diamond in driving ‘entry’ we construct an additional data set that identifies an entire set of scientists – those working in the field of

²⁹ Results are not affected if we also include researchers that moved shortly before Diamond opened. Note that we investigate movements within the UK only for the set of related authors because we do not have the complete set of any publications by all Diamond authors before the opening of Diamond.

³⁰ There are only 28 authors that moved between geographically distinct institutions after Diamond opened.

³¹ The lack of movement may be at least partly explained by the fact that (senior) academic positions are not easily changed, especially because locating in direct proximity to Diamond would most likely require a position either at Oxford University or the Rutherford Appleton Laboratory in Didcot, which are both very competitive workplaces.

crystallography – that can be thought of as ‘at risk’ of conducting research that might benefit from Diamond (either directly or indirectly). For this at-risk group, we can consider whether the likelihood of conducting work at Diamond depends on distance to Diamond.³² Crystallography is well-suited for our purposes. It existed as a field pre-Diamond. In fact, it emerged as a scientific field as a direct result of technological advances achieved through synchrotron radiation. The SRS, for example, is credited with pioneering the field of protein crystallography (STFC, 2010) and was more generally considered as world class in the field of crystallography. Moreover, scientists tend to specialise in crystallography, which makes it possible to identify a specific group of scientists for which synchrotron radiation plays a crucial research role. Crystallography also cuts across disciplines (most notably biology, (bio)chemistry and physics), which means considering this group helps us capture the effect across a range of scientific disciplines. Unfortunately, this also means that crystallographers can be found in different academic departments (and few universities have dedicated crystallography departments), which makes data collection labour intensive. Online Appendix B.6 discusses the data collection in more detail.

We start by analysing whether the distance to Diamond affects the location of crystallographers (i.e. the spatial distribution of the at-risk group) and the decision of crystallographers to use Diamond. Results are presented in Table 5 (for descriptive statistics see online Appendix Table E3). Column (1) presents results from estimating (5) with the TTWA-by-year count of crystallographers as the dependent variable. Results indicate that the number of crystallographers within 25 kilometres of Diamond increases significantly post-Diamond with no corresponding effect for Daresbury. As indicated in the last row of the Table, the difference between the coefficient on the Diamond and Daresbury 25 kilometres ring dummies is significant at the 1% level. Column (2) focuses specifically on those crystallographers that use Diamond by estimating (5) with the dependent variable defined as the TTWA-by-year count of scientists that eventually become Diamond authors.³³ Once again, we see significant positive effects within 25 kilometres of Diamond but no such effects for Daresbury, with the equality of the coefficients rejected at the 1% significance level. Finally, in column (3), we report results when we use as the dependent variable, the TTWA-by-year count of crystallographers that author related articles. That is, we consider whether proximity to Diamond also affects the probability of becoming a related article author. We again find significant positive effects of proximity to Diamond, with no such effect of proximity to Daresbury. Overall, these results suggest that the effect at the extensive margin is driven both by an increase in the number of crystallographers that are based close to Diamond and a positive effect of proximity to Diamond on the tendency for crystallographers to become authors of either Diamond or related articles.

The results in Table 5 implicitly equate the location of scientists with the location of the input that scientists provide. While this makes sense for physical production

³² This analysis is similar to Borjas and Doran (2012) who analyse how the choice of research field of American mathematicians reacts to the sudden influx of Soviet mathematicians following the collapse of the Soviet Union. However, our analysis focuses on the effects of an infrastructure rather than a supply shock.

³³ Since we can construct this count over the entire 2000–10 period given that we have the entire publication record for crystallographers, we can estimate the interaction of the ring dummies with the post-Diamond dummy.

Table 5
'Entry' of Crystallographers (209 TTWAs; 2000–10)

	No. of all crystallographers	No. of Diamond crystallographers	No. of related crystallographers
	(1)	(2)	(3)
Diamond $\times I(t \geq 2007)$			
25 kilometres	25.951*** (2.332)	20.643*** (1.673)	12.967*** (1.433)
125 kilometres	0.116 (0.207)	0.104 (0.146)	-0.002 (0.119)
175 kilometres	-0.116 (0.102)	-0.097 (0.069)	-0.071 (0.057)
Daresbury $\times I(t \geq 2007)$			
25 kilometres	-0.274 (0.458)	0.282 (0.292)	0.629 (0.412)
125 kilometres	-0.177 (0.170)	-0.032 (0.076)	-0.006 (0.067)
175 kilometres	-0.106 (0.165)	0.045 (0.149)	0.027 (0.098)
Time dummies	Yes	Yes	Yes
Covariates	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes
R ²	0.495	0.410	0.315
Observations	2,299	2,299	2,299
$\theta^{r=1} = \gamma_{DI}^{r=1} - \gamma_{DA}^{r=1}$	26.225*** (2.268)	20.361*** (1.569)	12.337*** (1.311)

Notes. Dependent variable in column (1) is count of all crystallographers by TTWA-year; dependent variable in column (2) is count of all crystallographers that are 'Diamond authors' by TTWA-year; dependent variable in column (3) is count of all crystallographers that are 'related authors' by TTWA-year. Robust standard errors clustered at the TTWA level. All regressions include a constant. Covariates include %NVQ4 and above, Labour force, No. of Wellcome grants, £'000 awarded by Wellcome, and £'000 awarded by STFC. * Significant at 10%, ** at 5%, *** at 1%.

functions (where, e.g. a machine's capital input cannot be separated from the machine's location) this is not necessarily true for knowledge production. In particular, co-authoring means that scientists' input in to knowledge production at both their own and their co-authors' location. Of course, for any one co-authoring relationship this effect works in both ways. However, spatial disparities in the amount of co-authoring could alter the distribution of effective inputs across locations. Specifically, if proximity to Diamond increases the number of co-authors, then this is another way in which increased inputs could explain the overall results. Because we have the full publication record of crystallographers, including the names of co-authors, we can use the crystallographers data to consider whether this 'indirect' input channel matters. To do this, we restrict our attention to the set of crystallographers that publish either a Diamond or related article and count the number of their co-authors by year. We then estimate (5) using this count of co-author by year at the individual crystallographer level. Results, reported in columns (1) and (2) of Table 6, show that Diamond and related article authors close to Diamond do, indeed, have more co-authors. However, this is also true of Diamond and related article authors close to Daresbury and we cannot reject equality of the two coefficients. If we are

Table 6
Co-authoring (Crystallographers; 2000–11)

	No. of co-authors		No. of new co-authors	
	Diamond	Related	Diamond	Related
	(1)	(2)	(3)	(4)
Diamond $\times I(t \geq 2007)$				
25 kilometres	0.957*** (0.320)	1.270*** (0.359)	0.470*** (0.162)	0.537*** (0.176)
125 kilometres	-0.306 (0.238)	-0.397 (0.268)	-0.089 (0.124)	-0.200 (0.128)
175 kilometres	0.429 (0.288)	0.467 (0.325)	-0.393 (0.229)	-0.318 (0.206)
Daresbury $\times I(t \geq 2007)$				
25 kilometres	1.650 (1.238)	1.853 (1.439)	0.971* (0.505)	1.102* (0.581)
125 kilometres	-0.446 (0.314)	-0.469 (0.393)	-0.149 (0.179)	-0.258 (0.221)
175 kilometres	-1.726*** (0.484)	-1.904*** (0.541)	-1.017*** (0.234)	-1.108*** (0.244)
Time dummies	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes
R ²	0.083	0.083	0.060	0.067
Observations	2,780	1,939	2,780	1,939
$\theta^{r=1} = \gamma_{DI}^{r=1} - \gamma_{DA}^{r=1}$	-0.692 (1.245)	-0.582 (1.449)	-0.500 (0.509)	-0.565 (0.590)

Notes. Dependent variable in columns (1) and (2) count of co-authors of a given crystallographer by author-year; in columns (3) and (4): count of new co-authors of a given crystallographer by author year. Robust standard errors clustered at the author level. All regressions include a constant. * Significant at 10%, ** at 5%, *** at 1%.

interested in the extensive margin, it might be more relevant to only consider new co-authoring relationships. When we do this, the results as reported in columns (3) and (4), are essentially unchanged. Indirect increases in inputs via asymmetries in co-authoring relationships do not explain the overall effects of proximity to Diamond.

Our results so far suggest that direct changes at the extensive margin (the number of scientists) explain our overall effects. We find no evidence that co-authoring provides an indirect input increase nor do our earlier results suggest that Diamond has a productivity effect that depends on distance to the facility. Before concluding, we look for one other indirect effect which is of interest – whether Diamond has any effect on the total productivity of researchers that produce Diamond related articles. To do this, we focus again on the set of crystallographers who have produced either a Diamond or related article. We then use the total publication count per crystallographer by year as the outcome variable. Table 7 shows results for both Diamond and related author counts. In column (1), we do not find any evidence for an increase in the number of publications among Diamond authors located within 25 kilometres of Diamond (or Daresbury). This contrasts with the positive publication effect that we found for Diamond articles as reported in Table 4,

Table 7
Total Output Effects for Crystallographers (Crystallographers; 2000–11)

	Diamond	Related articles				
	(1)	(2)	(3)	(4)	(5)	(6)
Diamond $\times I(t \geq 2007)$						
25 kilometres	0.803 (0.806)	1.442* (0.822)	1.596* (0.822)	1.430* (0.825)	1.448* (0.826)	1.430* (0.825)
125 kilometres	-1.332* (0.711)	-1.341* (0.706)	-1.313* (0.691)	-1.353* (0.709)	-1.336* (0.706)	-1.353* (0.709)
175 kilometres	0.041 (1.504)	0.013 (1.462)	0.138 (1.485)	0.007 (1.458)	0.007 (1.464)	0.007 (1.458)
Daresbury $\times I(t \geq 2007)$						
25 kilometres	2.346 (3.041)	3.016 (3.474)	3.166 (3.474)	3.003 (3.476)	3.019 (3.476)	3.003 (3.476)
125 kilometres	-0.641 (0.834)	-0.418 (0.886)	-0.400 (0.887)	-0.431 (0.888)	-0.419 (0.886)	-0.431 (0.888)
175 kilometres	-1.818* (1.066)	-2.143** (1.065)	-1.986* (1.064)	-2.157** (1.068)	-2.148** (1.067)	-2.157* (1.068)
Co-author $\times I(t \geq 2007)$			3.130*** (0.722)			
Citing author $\times I(t \geq 2007)$				1.898** (0.669)		
Unrelated journal $\times I(t \geq 2007)$					-0.207 (0.700)	
Citing author \times Unrelated journal $\times I(t \geq 2007)$						-1.898*** (0.669)
Time dummies	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.004	0.005	0.006	0.005	0.007	0.006
Observations	2,780	1,939	1,939	1,939	1,939	1,939
$\theta^{r=1} = \gamma_{DI}^{r=1} - \gamma_{DA}^{r=1}$	-1.542 (3.024)	-1.573 (3.490)	-1.570 (3.491)	-1.573 (3.491)	-1.570 (3.489)	-1.573 (3.491)

Notes. Dependent variable publication count by author-year. Robust standard errors clustered at the author level. All regressions include a constant. Unrelated journals are all journals other than the 118 journals in which 'Diamond authors' have published. * Significant at 10%, ** at 5%, *** at 1%.

suggesting that output of non-Diamond related articles decreases to offset increased Diamond article output (with the caveat that neither Diamond effect is significantly different from the corresponding Daresbury ring dummy). By contrast, results reported in column (2) do suggest an increase in the number of articles for authors of related articles within the 25 kilometres distance ring around Diamond. Note however that the significant positive effect for Diamond is smaller than the (insignificant) effect for Daresbury so we cannot rule out the possibility that this effect is driven by, say, the tendency of existing clusters of research to strengthen over time. Columns (3)–(6) add variables that capture whether:

- (i) a related author has co-authored with a Diamond author pre-2007;
- (ii) a related author has cited a Diamond author;
- (iii) whether a related author has published in a 'Diamond journal' (i.e. any journal in which a Diamond article is published); and
- (iv) the combination of (ii) and (iii).

We think of these variables as capturing a number of other ‘non-spatial’ mechanisms through which Diamond may affect productivity of crystallographers in terms of their total output. These results suggest that co-authors and citing authors experience a larger increase in publications post-Diamond whereas there is no effect for crystallographers that have published in Diamond relevant journals before Diamond opened. Overall, this set of results suggests that any effect on the intensive margin has no strong spatial dimension and is instead limited to the indirect effect of Diamond via co-author and citation links between related and Diamond authors.

7. Conclusion

Does the location of basic scientific research infrastructure affect its use and impact? This question is difficult to answer because the locations of scientific facilities are chosen in order to maximise their impact, posing a formidable challenge to empirical work that attempts to assess the causal relationship between location choice and impact. We address this question in the context of the Diamond Light Source, a 3rd generation synchrotron, in the UK and ask whether the location choice affected where scientific research – that benefits from the existence of Diamond – is conducted. The existence of a runner-up location (and a third geographical cluster) allows us to address partially the endogeneity inherent in the chosen location. Since the runner-up location in Daresbury was home to the previous 2nd generation synchrotron, we also account for possible distreatment effects. As an alternative to the runner-up location, we also construct a synthetic control location.

Overall, we find fairly strong evidence that Diamond caused the geographic concentration of relevant research close to Diamond (within a 25 kilometres radius) over and above what would have been expected had Diamond been located elsewhere. We show that changes in productivity do not explain the positive effects of proximity. We do, however, see an increase in inputs, specifically the number of scientists working on relevant research in proximity to Diamond – an effect that appears to be explained partly by relocation of scientists from abroad but mostly by the entry of new scientists. Our regressions using crystallographers suggest that this effect at the extensive margin is driven mostly by an increase in the number of scientists that are based close to Diamond. For knowledge production, we argue that co-authoring provides an additional mechanism for increasing inputs – although we show that this kind of indirect increase in inputs cannot explain our results. Interestingly, these effects are, if anything, stronger for related articles suggesting that Diamond created local externalities that benefit both Diamond-users and non-users working in related fields.

*Santa Clara University and Spatial Economics Research Centre (LSE)
London School of Economics, Spatial Economics Research Centre (LSE) and What Works Centre
for Local Economic Growth (LSE)*

Accepted: 16 October 2015

Additional Supporting Information may be found in the online version of this article:

Appendix A. Derivation of Equation (2).

Appendix B. Data Construction.

Appendix C. Relocation from Daresbury (SRS) to Oxford (Diamond).

Appendix D. Figures.

Appendix E. Tables.

Data S1.

References

- Abadie, A., Diamond, A. and Hainmueller, J. (2010). 'Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program', *Journal of the American Statistical Association*, vol. 105(490), pp. 493–505.
- Abadie, A., Diamond, A. and Hainmueller, J. (2015). 'Comparative politics and the synthetic control method', *American Journal of Political Science*, vol. 59(2), pp. 495–510.
- Abadie, A. and Gardeazabal, J. (2003). 'The economic costs of conflict: a case study of the basque country', *American Economic Review*, vol. 93(1), pp. 113–32.
- Ahlfeldt, G.M., Redding, S.J., Sturm, D.M. and Wolf, N. (2012). 'The economics of density: evidence from the Berlin Wall', *Econometrica*, vol. 83(6), pp. 2127–2189.
- Angrist, J.D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: an Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Audretsch, D.B. and Feldman, M.P. (1996a). 'R&D spillovers and the geography of innovation and production', *American Economic Review*, vol. 86(3), pp. 630–40.
- Audretsch, D.B. and Feldman, M.P. (1996b). 'Innovative clusters and the industry life-cycle', *Review of Industrial Organization*, vol. 11(2), pp. 253–73.
- Borjas, G.J. and Doran, K.B. (2012). 'Cognitive mobility: labor market responses to supply shocks in the space of ideas', Working Paper No. 18614. NBER.
- Conley, T.G. and Taber, C.R. (2011). 'Inference with 'difference in differences' with a small number of policy changes', *Review of Economics and Statistics*, vol. 93(1), pp. 113–25.
- David, P., Mowery, D. and Steinmueller, W.E. (1992). 'Analysing the economic payoffs from basic research', *Economics of innovation and New Technology*, vol. 2(1), pp. 73–90.
- Draca, M., Machin, S. and Witt, R. (2011). 'Panic on the streets of London: police, crime, and the July 2005 terror attacks', *American Economic Review*, vol. 101(5), pp. 2157–81.
- Feldman, M.P. and Francis, J. (2004). 'Homegrown solutions: fostering cluster formation', *Economic Development Quarterly*, vol. 18(2), pp. 127–37.
- Greenstone, M., Hornbeck, R. and Moretti, E. (2010). 'Identifying agglomeration spillovers: evidence from winners and losers of large plant openings', *Journal of Political Economy*, vol. 118(3), pp. 536–98.
- Hanson, G. (1997). 'Increasing returns, trade and the regional structure of wages', *ECONOMIC JOURNAL*, vol. 107(440), pp. 113–33.
- Jaffe, A. (1989). 'Real effects of academic research', *American Economic Review*, vol. 79(5), pp. 957–70.
- Jaffe, A., Trajtenberg, M. and Henderson, R. (1993). 'Geographic localization of knowledge spillovers as evidenced by patent citations', *Quarterly Journal of Economics*, vol. 108(3), pp. 577–98.
- Kantor, S. and Whalley, A. (2014a). 'Knowledge spillovers from research universities: Evidence from endowment value shocks', *Review of Economics and Statistics*, vol. 96(1), pp. 171–88.
- Kantor, S. and Whalley, A. (2014b). 'Research proximity and productivity: long-term evidence from agriculture', mimeo, Economics Department, University of California-Merced.
- Loder, N. (1999). 'Wellcome trust backs Rutherford to host synchrotron', *Nature*, vol. 402(6761), pp. 451–51.
- Malmendier, U., Moretti, E. and Peters, F.S. (2012). 'Winning by losing: evidence on the long-run effects of mergers', Working Paper 18024. NBER.
- STFC (2010). *New Light on Science: The Social & Economic Impact of the Daresbury Synchrotron Radiation Source (1981–2008)*, Swindon: Science & Technology Facilities Council.
- Wren, J.D., Kozak, K.Z., Johnson, K.R., Deakyn, S.J., Schilling, L.M. and Dellavalle, R.P. (2007). 'The write position', *EMBO Reports*, vol. 8(11), pp. 988–91.